

GI-19745307 264 --- **IVVDAF**PKNEKPIIYFKLYRQLPGEKEVAVDDAELN **QINS**
 ORF84 **WO 2006/078318** **IVVDAF**PKNEKPIIYFKLYRQLPGEKEVAVDDAELN **QINSEGO**PCT/US2005/027239
 GI-28810263 264 **IVVDAF**PKNEKPIIYFKLYRQLPGEKEVAVDDAELN **QINSEGO**
 GI-21909640 150 **IVVDAF**PKNEKPIIYFKLYRQLPGEKEVAVDDAELN **QINSEGO**
 GI-19224141 601 PHSS**IVVDAF**PKNEKPIIYFKLYRQLPGEKEVAVDDAELN **QINSEGO**PAKIQVQLLQNGQKMPNQIQEV

GI-19745307 305 --- **QEISVT**WTNQLVTDERGMAYIYSVKEVDK **NGELLEPKD** **YIKRED**
 ORF84 305 **QEISVT**WTNQLVTDERGMAYIYSVKEVDK **NGELLEPKD** **YIKRED**
 GI-28810263 305 **QEISVT**WTNQLVTDERGMAYIYSVKEVDK **NGELLEPKD** **YIKRED**
 GI-21909640 191 **QEISVT**WTNQLVTDERGMAYIYSVKEVDK **NGELLEPKD** **YIKRED**
 GI-19224141 661 TKD**QEISVT**WTNQLVTDERGMAYIYSVKEVDK **NGELLEPKD** **YIKRED**
IKDNSVYH**ERK**DLPKY**DA**NOEYKY**SVE**EVAVPDGYKVSYLGN**DI**FN**TR**ETEFV**EQ**NNF

GI-19745307 349 **GLTV**TNTYVKPTSG **HYDIEV**TFC **NGHIDIT**EDTTPDIVSGENQMK
 ORF84 349 **GLTV**TNTYVKPTSG **HYDIEV**TFC **NGHIDIT**EDTTPDIVSGENQMK
 GI-28810263 349 **GLTV**TNTYVKPTSG **HYDIEV**TFC **NGHIDIT**EDTTPDIVSGENQMK
 GI-21909640 235 **GLTV**TNTYVKPTSG **HYDIEV**TFC **NGHIDIT**EDTTPDIVSGENQMK
 GI-19224141 721 **NLEFG**NA**EL**EG**Q**SGSKII**DE**DTL**T**S**F**KGKKI**W**KND**T**AB**R**PA**I**Q**V**OL**Y**AD**G**V**E**GG**T**K

GI-19745307 394 **QIEG**EDS **KPIDEV**T **ENNLIE**FGKNTMPGEE
 ORF84 394 **QIEG**EDS **KPIDEV**T **ENNLIE**FGKNTMPGEE
 GI-28810263 394 **QIEG**EDS **KPIDEV**T **ENNLIE**FGKNTMPGEE
 GI-21909640 280 **QIEG**EDS **KPIDEV**T **ENNLIE**FGKNTMPGEE
 GI-19224141 781 **EL**SG**SG**NEWS**F**E**F**KN**L**K**K**Y**NG**T**G**ND**I**I**Y**S**V**KE**V**T**V**PT**G**Y**D**V**T**Y**S**AN**D**I**I**N**T**K**R**E**V**E**T**Q**G**

GI-19745307 424 **DGTNS**NKYEEVEDSRPVD**TLSGLS**SEQ**Q**SGDM**TIEEDS**ATH**IKFS**KRD
 ORF84 424 **DGTNS**NKYEEVEDSRPVD**TLSGLS**SEQ**Q**SGDM**TIEEDS**ATH**IKFS**KRD
 GI-28810263 424 **DGTNS**NKYEEVEDSRPVD**TLSGLS**SEQ**Q**SGDM**TIEEDS**ATH**IKFS**KRD
 GI-21909640 310 **DGTNS**NKYEEVEDSRPVD**TLSGLS**SEQ**Q**SGDM**TIEEDS**ATH**IKFS**KRD
 GI-19224141 841 **PKLEI**BET**LPL**ES**GAS**GG**TT**VEDSRPVD**TLSGLS**SEQ**Q**SGDM**TIEEDS**ATH**IKFS**KRD

GI-19745307 473 **IDGK**ELAGAT**MEL**RDS**SGNTI**ST**W**IS**DG**Q**V**K**D**F**Y**L**M**P**G**K**Y**T**F**V**E**T**A**A**P**D**G**Y**E**I**A**T**A**I**T**F**T**
 ORF84 473 **IDGK**ELAGAT**MEL**RDS**SGNTI**ST**W**IS**DG**Q**V**K**D**F**Y**L**M**P**G**K**Y**T**F**V**E**T**A**A**P**D**G**Y**E**I**A**T**A**I**T**F**T**
 GI-28810263 473 **IDGK**ELAGAT**MEL**RDS**SGNTI**ST**W**IS**DG**Q**V**K**D**F**Y**L**M**P**G**K**Y**T**F**V**E**T**A**A**P**D**G**Y**E**I**A**T**A**I**T**F**T**
 GI-21909640 359 **IDGK**ELAGAT**MEL**RDS**SGNTI**ST**W**IS**DG**Q**V**K**D**F**Y**L**M**P**G**K**Y**T**F**V**E**T**A**A**P**D**G**Y**E**I**A**T**A**I**T**F**T**
 GI-19224141 901 **IDGK**ELAGAT**MEL**RDS**SGNTI**ST**W**IS**DG**Q**V**K**D**F**Y**L**M**P**G**K**Y**T**F**V**E**T**A**A**P**D**G**Y**E**I**A**T**A**I**T**F**T**

GI-19745307 533 **VNEQ**Q**V**T**W**NG**K**AT**IG**DA**H**I**V**ND**A**Y**K**PT**KG**SG**Q**VID**IEE**KL**F**DE**Q**GH**S**SG**ST**TE**IED**SK**S**
 ORF84 533 **VNEQ**Q**V**T**W**NG**K**AT**IG**DA**H**I**V**ND**A**Y**K**PT**KG**SG**Q**VID**IEE**KL**F**DE**Q**GH**S**SG**ST**TE**IED**SK**S**
 GI-28810263 533 **VNEQ**Q**V**T**W**NG**K**AT**IG**DA**H**I**V**ND**A**Y**K**PT**KG**SG**Q**VID**IEE**KL**F**DE**Q**GH**S**SG**ST**TE**IED**SK**S**
 GI-21909640 419 **VNEQ**Q**V**T**W**NG**K**AT**IG**DA**H**I**V**ND**A**Y**K**PT**KG**SG**Q**VID**IEE**KL**F**DE**Q**GH**S**SG**ST**TE**IED**SK**S**
 GI-19224141 961 **VNEQ**Q**V**T**W**NG**K**AT**IG**DA**H**I**V**ND**A**Y**K**PT**KG**SG**Q**VID**IEE**KL**F**DE**Q**GH**S**SG**ST**TE**IED**SK**S**

GI-19745307 593 **SDVI**IG**Q**Q**G**
 ORF84 593 **SDVI**IG**Q**Q**G**
 GI-28810263 593 **SDVI**IG**Q**Q**G**EV**VD**TT**ED**T**Q**SG**MT**G**H**S
 GI-21909640 479 **SDVI**IG**Q**Q**G**EV**VD**TT**ED**T**Q**SG**MT**G**H**S**SG**ST**TE**IED**SK**S**SDVI**IG**Q**Q**G**EV**VD**TT**ED**T**Q**SG**MT**
 GI-19224141 1021 **SDLI**IG**Q**Q**G**EV**VD**TT**ED**T**Q**SG**MT**G**H**S

GI-19745307 602 **QI**V**ET**T**ED**T**Q**T**G**M**H**G**D**S**G**C**R**T**E**V**E**D**T**K**L**V**Q**S**F**H**F**D**N**K
 ORF84 602 **QI**V**ET**T**ED**T**Q**T**G**M**H**G**D**S**G**C**R**T**E**V**E**D**T**K**L**V**Q**S**F**H**F**D**N**K
 GI-28810263 619 **G**ST**T**K**I**ED**S**K**S**SD**V**I**V**GG**Q**Q**I**V**ET**T**ED**T**Q**T**G**M**H**G**D**S**G**C**R**T**E**V**E**D**T**K**L**V**Q**S**F**H**F**D**N**K
 GI-21909640 539 **G**H**S**SG**ST**K**I**ED**S**K**S**SD**V**I**V**GG**Q**Q**I**V**ET**T**ED**T**Q**T**G**M**H**G**D**S**G**C**R**T**E**V**E**D**T**K**L**V**Q**S**F**H**F**D**N**K
 GI-19224141 1047 **G**ST**T**K**I**ED**S**K**S**SD**V**I**V**GG**Q**Q**I**V**ET**T**ED**T**Q**T**G**M**H**G**D**S**G**C**R**T**E**V**E**D**T**K**L**V**Q**S**F**H**F**D**N**K

GI-19745307 639 **E**SE**S**N**S**E**I**PK**K**DK**PK**S**N**T**S**L**P**AT**G**ER**Q**H**N**M**F**F**I**M**V**T**S**C**S**L**I**S**S**V**F**V**I**S**L**K**T**K**R**L**S**S**C**
 ORF84 639 **E**SE**S**N**S**E**I**PK**K**DK**PK**S**N**T**S**L**P**AT**G**ER**Q**H**N**M**F**F**I**M**V**T**S**C**S**L**I**S**S**V**F**V**I**S**L**K**T**K**R**L**S**S**C**
 GI-28810263 676 **E**PE**S**N**S**E**I**PK**K**DK**PK**S**N**T**S**L**P**AT**G**ER**Q**H**N**M**F**F**I**M**V**T**S**C**S**L**I**S**S**V**F**V**I**S**L**K**T**K**R**L**S**S**C**
 GI-21909640 599 **E**PE**S**N**S**E**I**PK**K**DK**PK**S**N**T**S**L**P**AT**G**ER**Q**H**N**M**F**F**I**M**V**T**S**C**S**L**I**S**S**V**F**V**I**S**L**K**T**K**R**L**S**S**C**
 GI-19224141 1104 **E**PE**S**N**S**E**I**PK**K**DK**PK**S**N**T**S**L**P**AT**G**ER**Q**H**N**M**F**F**I**M**V**T**S**C**S**L**I**S**S**V**F**V**I**S**L**K**T**K**R**L**S**S**C**

FIGURE 54A

GI-19224WO 2006/078318-MRKNKLLLATAI LATALGTASLNQNVKRAETAGVSSGQLITKRSITPCT/US2005/027239

ORF80 PCT1 LBRKMRKNKLLLATAI LATALGTASLNQNVKRAETAGVVTICKSLQVTKMT-NDDEVLM
GI-21909636 1 MRKNKLLLATAI LATALGTASLNQNVKRAETAGVSENKLVKNTFDSYTDNEVLM
GI-28810259 1 MERERMRKNKLLLATAI LATALGTASLNQNVKRAETAGVSENKLVKNTFDSYTDNEVLM
GI-19745303 1 MRKNKLLLATAI LATALGTASLNQNVKRAETAGVSDGSILVWKKTFPSYTDNEVLM
GI-13621428 1 MRLRHLLTCAALTSFAAT-VHGETVNGAKLTVTNLDLVNS-ALH

GI-19224137 56 PNTDYTFVNFDSAAATCTESN-LPIKEGLAVN-NOEIKVSYSNITDKTSGNENQVVDVEMK
ORF80 60 PETAFTEFTEPDMTASCHREGS-LDIRNGIVBGLDKQVTVKYKNTDKTSQRTNLAQDFDSK
GI-21909636 56 PRADYTFKVEADSTASGKTKDGLKIKPGIVNGLT-EQILSYTNTDRPDSNVRSTEFDFSK
GI-28810259 61 PRADYTFKVEADSTASGKTKDGLKIKPGIVNGLT-EQILSYTNTDRPDSNVRSTEFDFSK
GI-19745303 56 PRADYTFKVEADSTASGKTKDGLDIRKPGVLDGLENTKTIHYVNSDKTTANERSVNFDFAN
GI-13621428 48 PNTDYTFKVEADSTASGKTKDGLDIRKPGVLDGLENTKTIHYVNSDKTTANERSVNFDFAN

GI-19224137 114 VTFPSVGTIYRYVVTENKCTAE-CVYDDTMMLDVYVGN--NEKCGLEPKYIVSKKCD SA
ORF80 119 VKFPAICVYRYMVSEKNDKED-CITYDTRKTVVDVYVGNKANNEECFEVLIVSKKCD SA
GI-21909636 115 VVFPFGIGVYRYTVSEKQGDVE-CITYDTRKTVVDVYVGN--KEGGGFEPKIVSKKCD SA
GI-28810259 120 VVFPFGIGVYRYTVSEKQGDVE-CITYDTRKTVVDVYVGN--KEGGGFEPKIVSKKCD SA
GI-19745303 116 VKFPGVGVYRYTVSEVNCNFA-GIAYDSQQHTVDVYVGN--REDGGFEAKYIVSKEGGQS
GI-13621428 103 VTFEKEVYVYKVTETEKIDRVPVSVSYDTTSYTVQVHVLWN-EEQQKPVATYIVGYKEGS-

GI-19224137 171 TRPPIQFVNSFETTSKLEKREVTGNTGDRKRAFHTLTLQFNEYEASSVVKTEENGO--
ORF80 178 TRKPIQFVNSIKTTSKLEKQITGNACDRKSENFHTLTLQFSEYKTKGSVVKTEENGO--
GI-21909636 172 VKRPIQFVNSFATTSKLVKKNVSCNTGELQKEDFTLTLNESTNEKNDQIVSLQKQNE--
GI-28810259 177 VKRPIQFVNSFATTSKLVKKNVSCNTGELQKEDFTLTLNESTNEKNDQIVSLQKQNE--
GI-19745303 173 DKRPIQFVNSFEDTTSKLVKKNVSCNTGELQKEDFTLTLNESTNEKNDQIVSLQKQNE--
GI-13621428 161 -KVPKQFVNSILDSTTLTKVKKVSGTGGDRSKDENFGLTLKANQYKASEKVMTEKTRKG

GI-19224137 229 ---TRDVKIGEAQKFTLNDSSVILSKLFPVGINVYKVEEAEANQCCYITTTATLNDG--EK
ORF80 236 ---KNDVIGTPIKFTLGHCKSVMLSKLPIGINVYKVEEAEANQCCYITTTATLNDG--EK
GI-21909636 230 ---KREVKIGTPIKFTLGHCKSVMLSKLPIGINVYKVEEAEANQCCYITTTATLNDG--EK
GI-28810259 235 ---KREVKIGTPIKFTLGHCKSVMLSKLPIGINVYKVEEAEANQCCYITTTATLNDG--EK
GI-19745303 231 ---TRDVKIGEAQKFTLNDSSVILSKLFPVGINVYKVEEAEANQCCYITTTATLNDG--EK
GI-13621428 220 QAPVQTEASIDQLVHFTLRDGESEIKVTLNLFVGVVAVVTEDDYKSEKVTINVEVSPQDGV

GI-19224137 283 LSTVNLG-QEHRNLDKTADEIVVTNKRDRTOVPTGVVGTLPAPFAVLSIVAIGGVYITRKN
ORF80 292 SDEHLSLTONOKTDESADDEIVVTNKRDRTOVPTGVVGTLPAPFAVLSIVAIGGVYITRKN
GI-21909636 285 SKVOLD-MEOKTDESADDEIVVTNKRDRTOVPTGVVGTLPAPFAVLSIVAIGGVYITRKN
GI-28810259 290 SKVOLD-MEOKTDESADDEIVVTNKRDRTOVPTGVVGTLPAPFAVLSIVAIGGVYITRKN
GI-19745303 285 TDCVNLG-DSKHTDKSHDEIVVTNKRDRTOVPTGVVGTLPAPFAVLSIVAIGGVYITRKN
GI-13621428 280 KNTAGNSTECELSTDKDMILHFTLNKNDREVPTGVAVTVAPYDALCIVAVGCALYVKNAN

GI-19224137 342 A
ORF80 352 A
GI-21909636 344 A
GI-28810259 349 A
GI-19745303 344 A
GI-13621428 340 A

FIGURE 55

173/487

GI-2190WO 2006/078318

PCT/US2005/027239

GI-28810261	1	MLFSVVMILTMLAFNOTVLAKDSTV
GI-19224139	1	MLFSVVMILTMLAFNOTVLAKDSTV
ORF82	1	MLFSVVMILTMLAFNOTVLAKDSTV
GI-19745305	1	MRKYWKMLFSVVMILTMLAFNOTVLAKDSTV

GI-21909638	20	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEEITTAGSGKASFSPLTFTTVGQYTYRV
GI-28810261	26	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEEITTAGSGKASFSPLTFTTVGQYTYRV
GI-19224139	26	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEEITTAGSGKASFSPLTFTTVGQYTYRV
ORF82	61	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEEITTAGSGKASFSPLTFTTVGQYTYRV
GI-19745305	32	QTSISVENVLERAGDSTPFPSIALESIDAMNTIEEITTAGSGKASFSPLTFTTVGQYTYRV

GI-21909638	80	YOKPSONKDYQADTTVFDVLVYVYDEDEGTLVAKVISRRAGDEEKSATFKPKRLVVKPIPI
GI-28810261	86	YOKPSONKDYQADTTVFDVLVYVYDEDEGTLVAKVISRRAGDEEKSATFKPKRLVVKPIPI
GI-19224139	86	YOKPSONKDYQADTTVFDVLVYVYDEDEGTLVAKVISRRAGDEEKSATFKPKRLVVKPIPI
ORF82	121	YOKPSONKDYQADTTVFDVLVYVYDEDEGTLVAKVISRRAGDEEKSATFKPKRLVVKPIPI
GI-19745305	92	YOKPSONKDYQADTTVFDVLVYVYDEDEGTLVAKVISRRAGDEEKSATFKPKRLVVKPIPI

GI-21909638	140	PRQPNI PKTPLPLAGEVKSLGILSIVLLGLLVLLYVKKLSRL
GI-28810261	146	PRQPNI PKTPLPLAGEVKSLGILSIVLLGLLVLLYVKKLSRL
GI-19224139	146	PRQPNI PKTPLPLAGEVKSLGILSIVLLGLLVLLYVKKLSRL
ORF82	181	PRQPNI PKTPLPLAGEVKSLGILSIVLLGLLVLLYVKKLSRL
GI-19745305	152	PRQPNI PKTPLPLAGEVKSLGILSIVLLGLLVLLYVKKLSRL

FIGURE 56

PCT/US05/27239 174/487

!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913505.pep from: 1 to: 1036 September 15, 2004 18:46

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences
z-scores computed from opt scores

z-score	obs (=)	exp (*)
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	5	5:====*
48	10	5:====*=====
50	12	4:====*=====
52	4	4:====*
54	6	3:====*
56	4	3:====*
58	5	2:====*
60	3	2:====*
62	0	1:*
64	1	1:*
66	1	1:*
68	0	1:*
70	1	1:*
72	0	0:
74	0	0:
76	1	0:*
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	1	0:*
98	0	0:

FIGURE 57

PCT/US05/27239

```

100 0 0:
102 0 0:
104 0 0:
106 1 0:=
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=

```

Joining threshold: 39, opt. threshold: 27, opt. width: 16, reg.-scaled

The best scores are:

initl initn opt z-sc E(54)..

```

/home/morama/gas/pili/align/gi-50913505.pep Begin: 1 End: 1036
! gi|50913505|ref|YP_059477.1| Collag... 6697 6697 6697 3452.1 9.6e-189
/home/morama/gas/pili/align/gi-19224141.pep Begin: 48 End: 144
! gi|19224141|gb|AAL86412.1|AF447492_... 63 100 159 105.9 0.023
/home/morama/gas/pili/align/gi-21909640.pep Begin: 147 End: 449
! gi|21909640|ref|NP_663908.1| protei... 35 35 136 96.3 0.08
/home/morama/gas/pili/align/gi-13621428.pep Begin: 57 End: 318
! gi|13621428|gb|AAK33238.1| hypothet... 33 33 91 75.6 1.1
/home/morama/gas/pili/align/gi-50913506.pep Begin: 33 End: 428
! gi|50913506|ref|YP_059478.1| Fimbri... 70 149 86 71.3 1.9
/home/morama/gas/pili/align/gi-13621432.pep Begin: 14 End: 56
! gi|13621432|gb|AAK33241.1| conserve... 40 65 78 68.0 2.9
/home/morama/gas/pili/align/gi-19745301.pep Begin: 241 End: 466
! gi|19745301|ref|NP_606437.1| putati... 52 52 73 64.8 4.3
/home/morama/gas/pili/align/gas15.pep Begin: 492 End: 739
! GAS15 GAS15 43 68 69 61.4 6.6
/home/morama/gas/pili/align/gi-21909636.pep Begin: 176 End: 298
! gi|21909636|ref|NP_663904.1| conser... 31 31 62 60.8 7.1
/home/morama/gas/pili/align/gi-28810259.pep Begin: 181 End: 303
! gi|28810259|dbj|BAC63197.1| hypothe... 31 31 62 60.7 7.2
/home/morama/gas/pili/align/gi-19224139.pep Begin: 90 End: 143
! gi|19224139|gb|AAL86410.1|AF447492_... 43 43 54 58.9 8.9
/home/morama/gas/pili/align/gi-19745305.pep Begin: 96 End: 149
! gi|19745305|ref|NP_606441.1| hypoth... 43 43 54 58.8 9
/home/morama/gas/pili/align/orf82.pep Begin: 125 End: 178
! TRANSLATE of: orf82.seq check: 4296... 43 43 54 58.2 9.6
/home/morama/gas/pili/align/gi-21909638.pep Begin: 84 End: 137
! gi|21909638|ref|NP_663906.1| hypoth... 43 43 52 58.0 9.9
\\End of List

```

```

gi-50913505.pep
/home/morama/gas/pili/align/gi-50913505.pep

```

```

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M
GAS10394]

```

```

SCORES Initl: 6697 Initn: 6697 Opt: 6697 z-score: 3452.1 E(): 9.6e-189
>>/home/morama/gas/pili/align/gi-50913505.pep (1036 aa)
initn: 6697 initl: 6697 opt: 6697 Z-score: 3452.1 expect(): 9.6e-189
Smith-Waterman score: 6697; 100.0% identity in 1036 aa overlap
(1-1036:1-1036)

```

```

10 20 30 40 50 60
gi-50913505. MYSRLKRELIVIVINRKKKYKLIRLMVTVGLIFSQVLVLPIRRRLGLQMISTQTKVIPQEIIVT

```

FIGURE 57A

PCT/US05/27239

```

670      680      690      700      710      720
gi-50913505. QEGSKLAIDFKARYPNLSIYSLGVSKDINSSTASSPVVLKYLSGEEHYGTTDTAELE
|||||
gi-50913505. QEGSKLAIDFKARYPNLSIYSLGVSKDINSSTASSPVVLKYLSGEEHYGTTDTAELE
670      680      690      700      710      720

730      740      750      760      770      780
gi-50913505. KTLNKIVEDSKLSQLGISDLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
|||||
gi-50913505. KTLNKIVEDSKLSQLGISDLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
730      740      750      760      770      780

790      800      810      820      830      840
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTSLFNVKASDEAYEKYKDNREGYS
|||||
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTSLFNVKASDEAYEKYKDNREGYS
790      800      810      820      830      840

850      860      870      880      890      900
gi-50913505. EMGDSDDTYGNTQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
|||||
gi-50913505. EMGDSDDTYGNTQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
850      860      870      880      890      900

910      920      930      940      950      960
gi-50913505. ADNNQKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGYLYETKAKLGYTLF
|||||
gi-50913505. ADNNQKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGYLYETKAKLGYTLF
910      920      930      940      950      960

970      980      990      1000     1010     1020
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTA
|||||
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTA
970      980      990      1000     1010     1020

1030
gi-50913505. TVALLFYRRQHRKKQY
|||||
gi-50913505. TVALLFYRRQHRKKQY
1030

```

gi-50913505.pep

/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]

```

SCORES  Init1: 63  Initn: 100  Opt: 159  z-score: 105.9  E(): 0.023
>>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)
initn: 100  init1: 63  opt: 159  Z-score: 105.9  expect(): 0.023
Smith-Waterman score: 159; 36.7% identity in 98 aa overlap
(895-990:48-144)

```

```

870      880      890      900      910      920
gi-50913505. SDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVDADNNQKLAGVEFELRKEDKK-IV
gi-19224141. FILGLLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKGTG
20      30      40      50      60      70

930      940      950      960      970      980

```

FIGURE 57C


```

1: |::|:
gi-21909640. DAYK-PTKGSQVIDIEEKLDPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGM
      450      460      470      480      490      500

```

```

gi-50913505.pep
/home/morama/gas/pili/align/gi-13621428.pep

```

```

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

```

```

SCORES  Init1: 33  Initn: 33  Opt: 91  z-score: 75.6  E(): 1.1
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
  initn: 33  init1: 33  opt: 91  Z-score: 75.6  expect(): 1.1
Smith-Waterman score: 95; 19.9% identity in 271 aa overlap
(568-819:57-318)

```

```

      540      550      560      570      580      590
gi-50913505. NKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTAL
      :|: : : :|: : :|: :|: :|: :|:
gi-13621428. VNGAKLIVTKNLDLVNSNALIPNTDFTFKLEPDTFVNEDGNKFKGVALNTPMTKVITYTNS
      30      40      50      60      70      80

```

```

      600      610      620      630      640      650
gi-50913505. HNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFFYFGEDGYR-----SGNGS
      ::|:| :|: : :|: | :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
gi-13621428. DKGGSNTKTAEFDFSEVTFEKPGVYYYKVFEEKIDKVPVGSYDPTTSYTVQVHVLWNEEQQ
      90      100     110     120     130     140

```

```

      660      670      680      690      700
gi-50913505. SNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTAS-SSPVVLKYLSGEEH
      : : : : :|:|:|: | :|:| :|: :|: | :|: :|: :|: :|: :|: :|: :|: :|: :|:
gi-13621428. KPVATYIVGYKEGSKVPI-QFKN---SLDSTTLIVKVKVSGTGGDRSKDFNFGLTLLKANQ
      150     160     170     180     190     200

```

```

      710      720      730      740      750
gi-50913505. YGITDTAELEKTLN----KIVEDSKLSQL---GISDLSQYVDYDKQPDVLT-----R
      || : : : :|:|:| : : : :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
gi-13621428. YYKASEKVMIEKTTKGGQAPVQTEASIDQLYHFTLKDGESIKVTNLPVGVVDYVVTEDDYK
      210     220     230     240     250     260

```

```

      760      770      780      790      800      810
gi-50913505. KSKVNDETEILYQKQVQVEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTL
      :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
gi-13621428. SEKVTINVEVSPQDCAVKNIAGNSTEQEETSTDKDMT-----ITFTNKKDFEVPTGVAMIV
      270     280     290     300     310

```

```

      820      830      840      850      860      870
gi-50913505. SENVKASDEAYEKYKDNENGRYSEMGDSDTDYGTNQTSSGKGLPSNSDASVNYMADGREQ
      :
gi-13621428. APYIALGIVAVGGALYFVKKKNA
      320     330     340

```

```

gi-50913505.pep
/home/morama/gas/pili/align/gi-50913506.pep

```

```

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes
MGAS10394].

```

```

SCORES  Init1: 70  Initn: 149  Opt: 86  z-score: 71.3  E(): 1.9
>>/home/morama/gas/pili/align/gi-50913506.pep (556 aa)

```

FIGURE 57E

initn: 149 initl: 70 opt: 86 Z-score: 71.3 expect(): 1.9
Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
(503-966:33-428)

```

      480      490      500      510      520
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHRYRGI VSVSDGIRRDDAVKNSLLGVNG-----L
      |:: || :: :||::: :: ::: |
gi-50913506. NRRETVREKILITAKKLM LACLAILAVVGLGMTRVS-ALSKDDTAQLKITNIEGGPTVTL
      10      20      30      40      50      60.

      530      540      550      560      570      580
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS
      : :: | | |:: | | : : | : : | : : | : |::
gi-50913506. YKIGEGVYNTNGDSFINFK-----YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSTEN
      70      80      90      100     110

      590      600      610      620      630      640
gi-50913505. LLDPNLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSG
      : | :: : |:: : ||| : ||| |:: : | || : :
gi-50913506. VSISNGTATYFNARGASVYIALLTGAT-----DGRTYNPILLAAS-----YNGEGNLVTK
      120     130     140     150     160

      650      660      670      680      690      700
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSPVLKYL SGE
      | : | :: | : | :: | |; |::: | : | : | : | : | : |
gi-50913506. NIDS--KSNLYGQ--TSVA---KSSLPSITKKVVTGTIDDV NKKTTSLGSLVLSYSLTFE
      170     180     190     200     210

      710      720      730      740      750      760
gi-50913505. EHYGITDTABLEKIFLNKIVEDSKLSQLGISDLSQYVDYDQKQPDV L VTRKSKVNDETE
      | | :::: || | : ||::| : : : : : : : | | : | : |
gi-50913506. LPSY-----TKEAVNKTVY-----VSDNMSEGLTF--NFNSLTVEWKGKMANITE
      220     230     240     250     260

      770      780      790      800      810      820
gi-50913505. ILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYT LSFNVKASDE
      | | ::: | | : | : | : | : : | : | : |
gi-50913506. -----DGSVMVENTKIGIAKEV NNGFNLSFIYDS--LESISPNI-----
      270     280     290

      830      840      850      860      870      880
gi-50913505. AYEKYKDNEGRYSEMCDSDTDYGINQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVI
      : | : : |::: | : | : : : | : | : | : : | : |
gi-50913506. SYKAVVNNKALVGEENPNKAEFFYSNNPTKGN TYDNLDRKPKD-KNGITSKEDSK-----
      300     310     320     330     340     350

      890      900      910      920      930      940
gi-50913505. QVKTVPITFTKVDADNNQKLAGVEFELRKEDKKIVWEKGT TGSNGQLNFKYLQKGKTY
      | | | : | ||| : : : | | : | : : | : | : |
gi-50913506. IVYTYQIAERKVDS-VSKTPLIGALFGVYDTSNKLI-DIVTINKNGYAI STQVSSGK-YK
      360     370     380     390     400

      950      960      970      980      990      1000
gi-50913505. LYETKAKLGYTLPENPWEVAVANNGDIKVKHP IEGELKSKDGSYMIKNYKIYQLPSSGGR
      : | || || : | : : | : :
gi-50913506. IKELKAPKGYSLNTTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGI
      410     420     430     440     450     460

```

gi-50913505.pep
/home/morama/gas/pili/align/gi-13621432.pep
gi|13621432|gb|AAK33241.1| conserved hypothetical protein [Streptococcus pyogene

FIGURE 57F

PCT/US05/27239 181/487

s)

SCORES Init1: 40 Initn: 65 Opt: 78 z-score: 68.0 E(): 2.9
 >>/home/morama/gas/pili/align/gi-13621432.pep (450 aa)
 initn: 65 init1: 40 opt: 78 z-score: 68.0 expect(): 2.9
 Smith-Waterman score: 78; 37.0% identity in 46 aa overlap
 (368-411:14-56)

```

      340      350      360      370      380      390
gi-50913505. KGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRLTPRSTRMK--RSPVVEKFECELE
                |  :::| |  ::|  |:|  :  :||  ||
gi-13621432.          MTRTNYQKKRMTCPVETEDITYRRKKIKGRRQAILAQFEPPELV
                10      20      30      40

      400      410      420      430      440      450
gi-50913505. HHKRIDYLGDNQNNPDTTIDDKEDHDSDLYRLYLDMTGKKNPLDIEVVVVDKSGSMQEG
                ||: |  ||: :  ||
gi-13621432. HHELI---GDSCTCPDCHGTLTEIGSVVQRQELVFIPAQLKRINHVVQHAYKQCQTCSDNSL
                50      60      70      80      90      100

```

gi-50913505.pep
 /home/morama/gas/pili/align/gi-19745301.pep

gi|19745301|ref|NP_606437.1| putative collagen binding protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 52 Initn: 52 Opt: 73 z-score: 64.8 E(): 4.3
 >>/home/morama/gas/pili/align/gi-19745301.pep (524 aa)
 initn: 52 init1: 52 opt: 73 z-score: 64.8 expect(): 4.3
 Smith-Waterman score: 95; 23.7% identity in 245 aa overlap
 (759-989:241-466)

```

      730      740      750      760      770      780
gi-50913505. DSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDTEILYQKDQVQVQEAG-KDIIDKVV
                ::::|||  :|  :|  :|  :|  :|  :|
gi-19745301. ETIDPDFNEGKEIKYTHILGADLFSYANNPRASFNDE--LLSQVKKVLEKGYRD--DSTT
                220      230      240      250      260

      790      800      810      820      830      840
gi-50913505. FTPKTTSPKPKVTLT---FKSDYKVVDD--EYTYLSEFNVKASDEAYEKYKDNENGRYSEM
                :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-19745301. YANLTSVEFRAATQLAIYYFTDSVDLDNLADYHGFALTFEALNATKEIVAYAEADRANLP
                270      280      290      300      310      320

      850      860      870      880      890
gi-50913505. GDSDDTY---GTNQTSS--GKGGLPNSDASVNYMADGREQKLPYKHPVIQVKTVPIPTFT
                :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-19745301. NISNLDYVVPNSNKYQSLIGTQYHP-ESLVDIIRMEDKQAPIIPITHKLTISKTVTGTI-
                330      340      350      360      370      380

      900      910      920      930      940      950
gi-50913505. KVDADNNQKLAGVEFELRKEDKKIVWEKGTIGSN-GQLNFKYLQKQK-TYYLYETKAKL
                ||  :||  :  :|:|:|  :  :  :||  :|  :|:  :  ||  :|  :  :  :
gi-19745301. ---AD--KKKEFNFEIHLKSSDQAI--SGTYPTNSGELT---VTDGKATFTLKDGESLI
                390      400      410      420      430

      960      970      980      990      1000      1010
gi-50913505. GYTLPEN-PWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNIYIQLPSSGGRGSQIFII
                ||  :  :|:|:|  :  :|  :|:  :  :|:  :  ||
gi-19745301. VEGLPSPGYSYEITETGASDYEVS--VNGK-NAPDGKATKASVKEDETITFENRKDLVPPT

```

FIGURE 57G

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.8 E(): 7.1
 >>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)
 initn: 31 init1: 31 opt: 62 Z-score: 60.8 expect(): 7.1
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
 (181-305:176-298)

```

      160      170      180      190      200      210
gi-50913505. DITTTLTFKNGLSLEGASTEANDPNVRVGVIVNPNNDTVQTTPTTIKQDADGKVKNLVFTGR
      || |:: | : :|:::|: ||:
gi-21909636. VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGN-----TGE
      150      160      170      180      190

      220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNIILKPKL
      | | : : : : | :|::| :||:: | : : : | : : : | : : | : : | : :
gi-21909636. LQKEFDFTLTLNESTNFKKQIVSLQK-GNEKFEVKIGTPYKFKLNKNGESIQLDKLPVGI
      200      210      220      230      240      250

      270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVLKTQYQSAKDSTPQTRDI
      : : ::: : || :| || :| ||: :| |
gi-21909636. TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESAD EIVVTNKRDTQVPTGV
      260      270      280      290      300      310

      330      340      350      360      370      380
gi-50913505. LFG EYTV EPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRLTFRSTRMKRS
gi-21909636. VGTLAPFAVLSIVAIGGVIIYITKRKKA
      320      330      340
  
```

gi-50913505.pep
 /home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.7 E(): 7.2
 >>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)
 initn: 31 init1: 31 opt: 62 Z-score: 60.7 expect(): 7.2
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
 (181-305:181-303)

```

      160      170      180      190      200      210
gi-50913505. DITTTLTFKNGLSLEGASTEANDPNVRVGVIVNPNNDTVQTTPTTIKQDADGKVKNLVFTGR
      || |:: | : :|:::|: ||:
gi-28810259. VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGN-----TGE
      160      170      180      190      200

      220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNIILKPKL
      | | : : : : | :|::| :||:: | : : : | : : : | : : | : : | : :
gi-28810259. LQKEFDFTLTLNESTNFKKQIVSLQK-GNEKFEVKIGTPYKFKLNKNGESIQLDKLPVGI
      210      220      230      240      250      260

      270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVLKTQYQSAKDSTPQTRDI
      : : ::: : || :| || :| ||: :| |
gi-28810259. TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESAD EIVVTNKRDTQVPTGV
      270      280      290      300      310      320
  
```

FIGURE 57I

PCT/US05/27239

gi-50913505. LFGEYTVPEPLVMNKGHNNTINIIYIRSTRPLGLKPIGAAPALIQPRSFRLTPRSTRMKRS
330 340 350 360 370 380

gi-28810259. VGTLPAPFAVLSIVAIGGVIIYITKRKKA
330 340

gi-50913505.pep
/home/morama/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.9 E(): 8.9
>>/home/morama/gas/pili/align/gi-19224139.pep (189 aa)
initn: 43 init1: 43 opt: 54 Z-score: 58.9 expect(): 8.9
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:90-143)

720 730 740 750 760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDLSQYVDYYDKQP--DVLVTRKSKVNDTEELL
|| || ||| : || |
gi-19224139. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLYV---TYDEDGTL
60 70 80 90 100 110
770 780 790 800 810 820
gi-50913505. YQKDOVQEQAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSEFNKASDEAY
| :|| : : :| || :|
gi-19224139. VAKVISRRAGDEEKSAITFKPKRLVKPIPPROPNIPKTPPLPLAGEVKSLLGILSIVLLGL
120 130 140 150 160 170

gi-50913505.pep
/home/morama/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP_606441.1| hypothetical protein [Streptococcus pyogenes MGAS82
32]

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.8 E(): 9
>>/home/morama/gas/pili/align/gi-19745305.pep (195 aa)
initn: 43 init1: 43 opt: 54 Z-score: 58.8 expect(): 9
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:96-149)

720 730 740 750 760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDLSQYVDYYDKQP--DVLVTRKSKVNDTEELL
|| || ||| : || |
gi-19745305. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLYV---TYDEDGTL
70 80 90 100 110 120
770 780 790 800 810 820
gi-50913505. YQKDOVQEQAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSEFNKASDEAY
| :|| : : :| || :|
gi-19745305. VAKVISRRAGDEEKSAITFKPKRLVKPIPPROPDIPKTPPLPLAGEVKSLLGILSIVLLGL
130 140 150 160 170 180

gi-50913505.pep
/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
generated symbols 1 to: 224.

FIGURE 57J

PCT/US05/27239
GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.2 E(): 9.6
>>/home/morama/gas/pili/align/orf82.pep (224 aa)
initn: 43 init1: 43 opt: 54 Z-score: 58.2 expect(): 9.6
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:125-178)

```

          720      730      740      750      760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDTEILL
          ||  ||      |||  :  ||  |
orf82.pep     ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
          100      110      120      130      140      150

          770      780      790      800      810      820
gi-50913505. YQKDQVQEQAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTTLNFKASDEAY
          |  ::||  :  :  ::|  ||  :|
orf82.pep     VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPPLPLAGEVKSLLGILSIVLLGL
          160      170      180      190      200      210
```

gi-50913505.pep
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 43 Initn: 43 Opt: 52 z-score: 58.0 E(): 9.9
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)
initn: 43 init1: 43 opt: 52 Z-score: 58.0 expect(): 9.9
Smith-Waterman score: 52; 31.6% identity in 57 aa overlap
(742-796:84-137)

```

          720      730      740      750      760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDTEILL
          ||  ||      |||  :  ||  |
gi-21909638. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
          60      70      80      90      100      110

          770      780      790      800      810      820
gi-50913505. YQKDQVQEQAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTTLNFKASDEAY
          |  ::||  :  :  ::|  ||  :|
gi-21909638. VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPPLPLAGEVKSLLGILSIVLLGL
          120      130      140      150      160      170
```

! Distributed over 1 thread.
! Start time: Wed Sep 15 18:46:18 2004
! Completion time: Wed Sep 15 18:46:21 2004

! CPU time used:
! Database scan: 0:00:00.1
! Post-scan processing: 0:00:02.2
! Total CPU time: 0:00:02.3
! Output File: gi-50913505.fasta

FIGURE 57K

!!SEQUENCE_LIST 1.0 PCT/US05/27239

(Peptide) FASTA of: gi-50913506.pep from: 1 to: 556 September 15, 2004 18:45

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences z-scores computed from opt scores

z-score	obs (=)	exp (*)
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	1	5:= *
48	8	5:====*====
50	24	5:====*=====
52	2	4:== *
54	3	3:==*
56	1	3:= *
58	2	2:==*
60	3	2:==*
62	0	2: *
64	2	1:*=
66	2	1:*=
68	1	1:*=
70	2	1:*=
72	0	0:
74	3	0:===
76	1	0:=
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:

FIGURE 58


```

100      0      0:
102      0      0:
104      0      0:
106      0      0:
108      0      0:
110      0      0:
112      0      0:
114      0      0:
116      0      0:
118      0      0:
>120     1      0:=
    
```

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are: initl initn opt z-sc E(55)..

```

/home/morama/gas/pili/align/gi-50913506.pep    Begin: 1   End: 556
! gi|50913506|ref|YP_059478.1| Fimbri... 3454 3454 3454 1016.2 4.7e-53
/home/morama/gas/pili/align/orf84.pep        Begin: 316  End: 567
! TRANSLATE of: orf84.seq check: 7868...   57   83  135   75.1   1.2
/home/morama/gas/pili/align/gi-19745307.pep   Begin: 316  End: 567
! gi|19745307|ref|NP_606443.1| protei...  57   83  135   75.1   1.2
/home/morama/gas/pili/align/gi-21909640.pep   Begin: 202  End: 524
! gi|21909640|ref|NP_663908.1| protei...  56   81  134   75.0   1.2
/home/morama/gas/pili/align/gi-28810263.pep   Begin: 316  End: 638
! gi|28810263|dbj|BAC63201.1| protein...  56   82  134   74.7   1.3
/home/morama/gas/pili/align/orf80.pep        Begin: 49   End: 352
! TRANSLATE of: orf80.seq check: 9824...  45   69  113   70.8   2.1
/home/morama/gas/pili/align/gi-19224137.pep   Begin: 25   End: 342
! gi|19224137|gb|AAL86408.1|AF447492...  45   69  109   69.8   2.4
/home/morama/gas/pili/align/gi-19224141.pep   Begin: 277  End: 645
! gi|19224141|gb|AAL86412.1|AF447492...  73   73  118   68.9   2.7
/home/morama/gas/pili/align/gi-21909636.pep   Begin: 44   End: 344
! gi|21909636|ref|NP_663904.1| conser...  45   98   96   66.1   3.8
/home/morama/gas/pili/align/gi-28810259.pep   Begin: 49   End: 349
! gi|28810259|dbj|BAC63197.1| hypothe...  45   98   96   66.0   3.8
/home/morama/gas/pili/align/gas15.pep        Begin: 222  End: 470
! GAS15 GAS15                               42   68   96   63.8   5
/home/morama/gas/pili/align/gi-13621428.pep   Begin: 17   End: 340
! gi|13621428|gb|AAK33238.1| hypothet...  41   41   87   63.6   5.2
/home/morama/gas/pili/align/gi-19224135.pep   Begin: 193  End: 462
! gi|19224135|gb|AAL86406.1|AF447492...  41   41   86   61.0   7
/home/morama/gas/pili/align/gi-50913505.pep   Begin: 503  End: 966
! gi|50913505|ref|YP_059477.1| Collag...  70   149  86   60.1   7.8
/home/morama/gas/pili/align/gi-13621430.pep   Begin: 60   End: 143
! gi|13621430|gb|AAK33240.1| hypothet...  43   67   67   59.2   8.7
/home/morama/gas/pili/align/gi-19745303.pep   Begin: 44   End: 344
! gi|19745303|ref|NP_606439.1| hypoth...  51   106  69   58.4   9.5
\\End of List
    
```

gi-50913506.pep

/home/morama/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

```

SCORES    Initl: 3454   Initn: 3454   Opt: 3454   z-score: 1016.2 E(): 4.7e-53
>>/home/morama/gas/pili/align/gi-50913506.pep                               (556 aa)
   initn: 3454   initl: 3454   opt: 3454   Z-score: 1016.2   expect(): 4.7e-53
Smith-Waterman score: 3454;   100.0% identity in 556 aa overlap
    
```

FIGURE 58A

```

      10      20      30      40      50      60
gi-50913506. MTNRRETVREKILITAKKMLACLAILAVVGLGMTRVSALSKDDTAQLKITNIEGGPTVT
|||||
gi-50913506. MTNRRETVREKILITAKKMLACLAILAVVGLGMTRVSALSKDDTAQLKITNIEGGPTVT
      10      20      30      40      50      60

      70      80      90     100     110     120
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSIS
|||||
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSIS
      70      80      90     100     110     120

      130     140     150     160     170     180
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYNGEENLVTKNIDSKSNLYGQTSV
|||||
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYNGEENLVTKNIDSKSNLYGQTSV
      130     140     150     160     170     180

      190     200     210     220     230     240
gi-50913506. AKSSLPSITKKVGTGTTIDDVNKKTTSLGSLVLSYSLTFELPSYTKAEVAVNKTIVVSDNMSEGL
|||||
gi-50913506. AKSSLPSITKKVGTGTTIDDVNKKTTSLGSLVLSYSLTFELPSYTKAEVAVNKTIVVSDNMSEGL
      190     200     210     220     230     240

      250     260     270     280     290     300
gi-50913506. TFFNFNSLTFVEWKGMANITEDGSVMVENTKIGIAKEVAVNNGFNLSFIYDSLESISPNIYSK
|||||
gi-50913506. TFFNFNSLTFVEWKGMANITEDGSVMVENTKIGIAKEVAVNNGFNLSFIYDSLESISPNIYSK
      250     260     270     280     290     300

      310     320     330     340     350     360
gi-50913506. AVVNNKAIVGEEGNPNKAEFFYSNNPTKGN TYDNLDKKPKDKNGITSKEDSKIVYTYQIA
|||||
gi-50913506. AVVNNKAIVGEEGNPNKAEFFYSNNPTKGN TYDNLDKKPKDKNGITSKEDSKIVYTYQIA
      310     320     330     340     350     360

      370     380     390     400     410     420
gi-50913506. FRKVDVSKTPLIGALFVGYDTSNKLIDIVTFNKNKYAISTQVSSGKYKIKELKAPKGY
|||||
gi-50913506. FRKVDVSKTPLIGALFVGYDTSNKLIDIVTFNKNKYAISTQVSSGKYKIKELKAPKGY
      370     380     390     400     410     420

      430     440     450     460     470     480
gi-50913506. LNTEFYELTANWVTAIVKTSANSKSTTYTSDKNKATDNSEQVGLKNGIFYSIDSRPTGN
|||||
gi-50913506. LNTEFYELTANWVTAIVKTSANSKSTTYTSDKNKATDNSEQVGLKNGIFYSIDSRPTGN
      430     440     450     460     470     480

      490     500     510     520     530     540
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNPKLGELPSTGSGTYLFFKAIGSA
|||||
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNPKLGELPSTGSGTYLFFKAIGSA
      490     500     510     520     530     540

      550
gi-50913506. AMIGAIGIYIVKRRKA
|||||
gi-50913506. AMIGAIGIYIVKRRKA
      550

```

FIGURE 58B

189/487

PCT/US05/27239
gi-50913506.pep
/home/morama/gas/pili/align/orf84.pep

TRANSLATE of: orf84.seq check: 7868 from: 1 to: 2088
generated symbols 1 to: 696.
GETSEQ from morama, September 13, 2004 17:07.

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
>>/home/morama/gas/pili/align/orf84.pep (696 aa)
initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
(232-462:316-567)

```

                210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKAEVANKTVVYVSDNMSEGLTFNFNSLTVVEWKGKMAN----
                |::| :|::: :: |::|:::
orf84.pep     EKEVAVDDAELKQINSEGGQQLSVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
                290      300      310      320      330      340

```

```

                260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGNFNSFIY-----DSLESISPNISYKAVVNNKAIKIVGEE
                |::: ::| || : |::: ::: : | :|::|::| : : | | ||:
orf84.pep     IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
                350      360      370      380      390

```

```

                320      330      340      350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIIVTYTQ
                ::| | || :: | |::| |::: |::| | :|::|::: :: :| :
orf84.pep     SKPIDEVTENNLIIEFGKNIMPGEEDGTNSNKYEEVEDSRPVDTLGSLSSQQGQSGDMTIE
                400      410      420      430      440      450

```

```

                360      370      380      390      400      410
gi-50913506. -----IAFRKVDVSKRTEPLIGAIFGVYDTSNKLIDIVTPTNKNGYAISTQVSSGKYIKKE
                | | | | :: | | | : : |::| | : | :| : : : | | | : |
orf84.pep     EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMGKPTFVE
                460      470      480      490      500      510

```

```

                420      430      440      450      460      470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGLKNGIF
                || || : | | : :: : ||:: |::: :: ::: | | | : | ||
orf84.pep     TAAPDGYEIAATAITFTVNEQQQVTVNGKATKGDALHVMV-DAYKPTKGSQGVIEDIEEKLP
                520      530      540      550      560      570

```

```

                480      490      500      510      520      530
gi-50913506. YSIDSRPITGNDVKEAYIESTKALTDGTFESKSNESGTFVLEETDIPNTRLGEELPSTGSGIG
orf84.pep     DEQCHSGSTTEIEDSKSSDVLIGGQGVIVETTEDTQTMHGDSGCKTEVEDTKLVQSFHF
                580      590      600      610      620      630

```

gi-50913506.pep
/home/morama/gas/pili/align/gi-19745307.pep

gi|19745307|ref|NP_606443.1| protein F2-like protein [Streptococcus pyogenes MGA S8232]

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
>>/home/morama/gas/pili/align/gi-19745307.pep (696 aa)
initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
(232-462:316-567)

FIGURE 58C

PCT/US05/27239

```

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGITSKEDSKIVYTYQ
          ::|          | || :: |          |::| |::: |::| | :|::|:: :: :| :
gi-21909640. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQQQSGDMTIE
          290          300          310          320          330          340

          360          370          380          390          400          410
gi-50913506. -----IAFRKVDVSKTPLIGALFVYDTSNKLIDIVTTNKNGYAIISTQVSSGKYKIKE
          | | | | :: | | | : : |::| | : | : : | : : : | | | : |
gi-21909640. EDSATHIKFSKRD-IDGKELAGATMELRDSGKTIS--TWISDGQVKDFYLMGKYTFVE
          350          360          370          380          390          400

          420          430          440          450          460          470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGLKNGLIF
          || || : | | : : : ||:: |:: : : : :: | | | : | ||
gi-21909640. TAAPDGYEVATAITFTVNEQQQVTVNGKATKGDHIVMV-DAYKPTKSGSQV-----
          410          420          430          440          450

          480          490          500          510          520          530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNESGTVLLETDIPNTKLGELPSTGSIG
          || : | : : : || : | : : : | : | | : : | : | : ||
gi-21909640. --IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-
          460          470          480          490          500

          540          550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
          | : : : | : : | : |
gi-21909640. TEIEDSKSSDVIIGGQGEVVDTEEDTQSGMTGHSGSTTKIEDSKSSDVIIVGGQGVIVETT
          510          520          530          540          550          560

```

gi-50913506.pep /home/morama/gas/pili/align/gi-28810263.pep

gi|28810263|dbj|BAC63201.1| protein F2-like protein [Streptococcus pyogenes SSI-1]

SCORES Initl: 56 Inits: 82 Opt: 134 z-score: 74.7 E(): 1.3
 >>/home/morama/gas/pili/align/gi-28810263.pep (733 aa)
 inits: 82 initl: 56 opt: 134 z-score: 74.7 expect(): 1.3
 Smith-Waterman score: 155; 23.6% identity in 347 aa overlap
 (232-547:316-638)

```

          210          220          230          240          250
gi-50913506. KPTSLGSLVLSYSLTTELEPSYTKKAVNKTYVSDNMSEGLTINENSLTVEWKGKMAN-----
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
gi-28810263. EKEVAVDDBELKQINSEGGQEI SVIWNQLVDE--KGMAYIYSVKEVDKNGELLEPKDY
          290          300          310          320          330          340

          260          270          280          290          300          310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGNFNLFIY-----DSLSEISPNISYKAVVNNKAVGEE
          | : : : | || : | : : : : : | | : | : | : : | : | | | |
gi-28810263. IKKEDGLTINTFYV--KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED-
          350          360          370          380          390

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGITSKEDSKIVYTYQ
          ::|          | || :: |          |::| |::: |::| | :|::|:: :: :| :
gi-28810263. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQQQSGDMTIE
          400          410          420          430          440          450

          360          370          380          390          400          410

```

FIGURE 58E

PCT/0505/27239

```

gi-50913506. -----IAFRKVDVSVKTPPLIGAIFGVYDTSNKLIDIVTINKNGYAISTQVSSGKYKIKE
| | | | :: | | : : | : | | : | : : : : | | : |
gi-28810263. EDSATHIKFSKRD-IDGKELAGATMELRDSGKTIS--TWISDGOVKDFYLMPGKYTFVE
460 470 480 490 500 510

420 430 440 450 460 470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSA NSKSTTYTSDKNKATDNSEQVGLKNGIF
| | | : | | : : : | | : | : : : : | | : | |
gi-28810263. TAAPDGYEVATAITFTVNEQQQVTVNGKATPKGDAHIVMV-DAYKPTKGSQV-----
520 530 540 550 560

480 490 500 510 520 530
gi-50913506. YSIDS RPTGNDVKEAYIESTKALTDGTTFSKSNESGTVLLETDIPNTKLGELPSTGSGIG
| | : | : : : | | : | : : : | : | : | : : | : |
gi-28810263. --IDIEKLPD-EQGHSGSTTEIEDSKSSDVLIGGQGEVVDTTE--DTQSGMTGHSGST-
570 580 590 600 610 620

540 550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
| : : : | : : : |
gi-28810263. TKIEDSKSSDVIVGGQGI VETTEDTQGMHGDSGRKTEVEDTKLVQSFHFDNKEPESNS
630 640 650 660 670 680

```

gi-50913506.pep
/home/morama/gas/pili/align/orf80.pep

TRANSLATE of: orf80.seq check: 9824 from: 1 to: 1056
generated symbols 1 to: 352.
GETSEQ from morama, September 13, 2004 17:11.

SCORES Init1: 45 Initn: 69 Opt: 113 z-score: 70.8 E(): 2.1
>>/home/morama/gas/pili/align/orf80.pep (352 aa)
initn: 69 init1: 45 opt: 113 Z-score: 70.8 expect(): 2.1
Smith-Waterman score: 123; 22.8% identity in 311 aa overlap
(284-556:49-352)

```

260 270 280 290 300 310
gi-50913506. KMANITEDGSVMVENTKIGIATAKEVNNNGFNLSFIYDSLESISPNLSYKAVVN-NKAIVGEE
:: | | : | : | : : : : : : | |
orf80.pep. ATALGTASLNQNVKAETAGVVTGKSLQVTKMTYDDEEVIMPETAFTFTTEPDMTASCKE
20 30 40 50 60 70

320 330 340 350 360 370
gi-50913506. GNPN-KAEFFYSNNPTKGN TYDNLDRKDPKNGCITSKEDSKIVYTYQIAFRKVDVSVKTP
| : : | : : : | | | : | : : : | | : : : | : :
orf80.pep. GSLDIKNGIVEGLDKQVTVKVKYKNTDRPSQTK-LAOFDFSKVKFPAIGVYRYMVSEKNDK
80 90 100 110 120 130

380 390 400 410 420
gi-50913506. LIGAI FGVYDTSNKLIDIVTINKNG-----YAISTQ-VSSGKYKIKELKAPKGYSLN
| : | | : : : | : : | : | : | : : | : : | : |
orf80.pep. KDGI---TYDDKKWTVDVYVGNKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLK
140 150 160 170 180 190

430 440 450 460
gi-50913506. TETYEITANW-----VTATVKTSA NSKSTTYTSDKNKATDNSEQV-----WLKNGI
| : | | : | | : | : : : : : : : : | : |
orf80.pep. IEK-QITGNAGDRKKSFNFTLTLQPSYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGK
200 210 220 230 240 250

470 480 490 500 510

```

FIGURE 58F

gi-50913506.pep
/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.1 E(): 3.8
>>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)
initn: 98 init1: 45 opt: 96 Z-score: 66.1 expect(): 3.8
Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
(298-556:44-344)

```

      270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNNGFNLSFIYDSLESISPNI SYKAVVNNKAI VGEEGNPNKAEFFYSNNPT
      :: : ::|::: : | | ::: :
gi-21909636. ATALGTASLNQN VKAETAGVSENAKLI VKKTFDSYTDNEV LMPKADYTFKVE---ADSTA
      20      30      40      50      60      70

      330      340      350      360
gi-50913506. KGNTYDNL DKKPDKNGIT-----SKEDSKIVYTYQ IAFRKV-----DSVSK
      :|:| |:|: ||:| :| |||: | :| || :|:|
gi-21909636. SGKTKDGL EIKPGIVNGLTEQ IISYTNIDKPD SKVKST-EFDFSKV VFPGIGVYRYTVSE
      80      90      100      110      120

      370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVTTNKNGYALSTQVSSGKYKIKELKAPKGY--SINTETY
      | : | : |||: :|: : ||:| : : : | : : :| | : : | :| :
gi-21909636. KQ--GDVEGITYDTPKKTVDVYVGNKEGGGFEPK FIVSKEQGT DVKKPVNFNNSFATTSL
      130      140      150      160      170      180

      430      440      450      460
gi-50913506. EITANVWTATV KTSAN-----SKSTTYTSDK---NKATDNSE-QVGV-----LKNGI
      :: | | : : : : :|:|:| :|: :|:|:| :|:| :|:|
gi-21909636. KVKKNVSGN TIGELQKEFDFTL TLNESTNFKK DQIVSLQKGN EKFVKGITPYKFKLKNGE
      190      200      210      220      230      240

      470      480      490      500      510
gi-50913506. FYSIDSRPTGNDV KEAYIESTKALTDGTTFSKS-NEGSGTVLL-----ETD-----
      ::|: |:| | :|:| || :| :|:|:| : : :|
gi-21909636. SIQLDKLPV GITYKVNEM EANK---DGYKTTAS LKEGDGQSKMY QLDMEQKTD EADEIV
      250      260      270      280      290      300

      520      530      540      550
gi-50913506. IPNTIKGELPSTG SIGTYLEKAIGSAAMIGATG IYIVKRRKA
      : | : : :| || :|| :| :| :|:| :|:| :|:|
gi-21909636. VTNKRDTQVP-TGVVGT LAPFAVL SIVAIGGV-IYIVKRRKA
      310      320      330      340

```

gi-50913506.pep
/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.0 E(): 3.8
>>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)
initn: 98 init1: 45 opt: 96 Z-score: 66.0 expect(): 3.8
Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
(298-556:49-349)

FIGURE 58I

```

270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
                :: : ::|::: :   | :   :: :
gi-28810259. ATALGTASLNQNVKAETAGVSENAKLIVKKTFFDSYTDNEVLMPKADYTFKVE---ADSTA
20        30        40        50        60        70

330      340      350      360
gi-50913506. KGNTYDNLDKKPKDKNGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK
:|:| |:|: ||  ||:|   :| |||: | : : | ||   :||:
gi-28810259. SGKTKDGLEIKPGIVNGLTEQIISYTNIDKPDSPKVKST-EFDFSKVVFPGIGVYRYTVSE
      80      90      100      110      120      130

370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVITNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY
      | : | : |||:: : | : : ||:| : : : : | : : :| | : : | : | :
gi-28810259. KQ--GDVEGITYDTKKWTFVDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSL
      140     150     160     170     180     190

430      440      450      460
gi-50913506. EITANVWTATVKTSAN-----SKSTTYTSDK---NKATDNSE-QVGW-----LKNGI
: : | | : : : : :| : : : :| : : : | : : | |||
gi-28810259. KVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQKNEKFEVKIGTPYKFKLKNGE
      200     210     220     230     240     250

470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYLESTKALTDGTTESKS-NEGSGTVLL-----ETD-----
: : | : | : | :| : : :| || : | :| :| : : : :| :|
gi-28810259. SIQLDKLPVGIYKVNEMEANK---DGYKTTASLKEGDGQSKMYQLDMEQKTDSEADIEIV
      260     270     280     290     300

520      530      540      550
gi-50913506. IPNTKLGELPSTGSIGTYLFAIGSAAMIGAIGIYIVKRRKA
: | : :| || :|| | : | : ||: : ||:| :| :|
gi-28810259. VTKRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITKRRKA
310      320      330      340

```

gi-50913506.pep
/home/morama/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Init1: 42 Initn: 68 Opt: 96 z-score: 63.8 E(): 5
>>/home/morama/gas/pili/align/gas15.pep (762 aa)
initn: 68 init1: 42 opt: 96 Z-score: 63.8 expect(): 5
Smith-Waterman score: 96; 23.4% identity in 269 aa overlap
(283-535:222-470)

```

260      270      280      290      300
gi-50913506. GKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLES-ISPNISYKA--VVNN--K
                ||:: :|::: |:||: : | : :
gas15.pep     VWYYS DNAPISNPDESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPKQVPDDFQL
      200     210     220     230     240     250

310      320      330      340      350      360
gi-50913506. AIVGEEGNPNKAEFFYSNNPTKGNTRYDNLDKKPKDKNGIT'SKEDSKIVYTYQIAFRK--V
:| | : :| : | :| : | : | | | : : : | : :| :
gas15.pep     SIFESEDKGDKYNKGYQNLLSGGLVPT---KPPTPGDPMPNPQPQ---TTSVLIRKYAI
      260     270     280     290     300

370      380      390      400      410      420

```

FIGURE 58J

PCT/US05/27239

```

gi-50913506. DSVSKTPLIGAIFGVY-DTSNKL-IDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYSLN
: || | || : : | : | : : | | | : : | | : || : | || :
gas15.pep    GDYSKL-LEGATLQLTGDNVNSFQARVFSNDIGERI--ELSDGTYTLTELNSPAGYSIA
          310          320          330          340          350          360

```

```

          430          440          450          460          470
gi-50913506. TE-TYEITANWVTATV--KTSANSKSTTYTSDKNKATDNSEQVGLKN---GIFYSIDSR
: : : | : | : : | | : : : : | : | : : | | : : | | :
gas15.pep    EPITFKVEAGKVYTIIDGKQIENPNKEIVEPYSVEAYNDFEEFVLTTONYAKFYAKNK
          370          380          390          400          410          420

```

```

          480          490          500          510          520          530
gi-50913506. PTGNDVKEAYIESTKALTDGTTFSKSNESGTVLLETDPINPKLGLPSTGSGITYLFKA
: : : | : : | : | | : : | : | : | : | | | |
gas15.pep    NGSSQVVYCFNADLKSPD-----SEDGKTKMT-----PDFTTGEVKYTHIAGRDLFKY
          430          440          450          460          470

```

```

          540          550
gi-50913506. IGSAAMIGAIGIYIVKRRKA
gas15.pep    TVKPRDTPDPTFLKHKKVIEKGYREKQQAIEYSGLTETQLRAATQLAIYYFTDSAELDK
          480          490          500          510          520          530

```

```

gi-50913506.pep
/home/morama/gas/pili/align/gi-13621428.pep

```

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

```

SCORES  Init1: 41  Initn: 41  Opt: 87  z-score: 63.6  E(): 5.2
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
initn: 41 init1: 41 opt: 87 Z-score: 63.6 expect(): 5.2
Smith-Waterman score: 109; 22.6% identity in 345 aa overlap
(256-556:17-340)

```

```

          230          240          250          260          270          280
gi-50913506. VNKTVYVSDNMSEGLTFNFNSLTVIEWKGMANTTEDGSMVENTKIGIAKEVNNNGFNLSF
          | | | : : : : : | | :
gi-13621428. MKLRHLLLGAALTSAATTVHGETVVGAKLTVTK-----NLDL
          10          20          30          40

```

```

          290          300          310          320          330          340
gi-50913506. TYDSLESISPNISYKAVVNNKATVGEENPNKAEFFYSNNPTKGNLYDNLDKKPKDRKNGI
: : : | : : | : : | : : | : : | : : | : : | : : | : : |
gi-13621428. VNSN--ALIPNTDETEKLEEDTIVNEDGNKFKGVAL--NIPMTKVTVYVNSDK--GGSNPK
          50          60          70          80          90

```

```

          350          360          370
gi-50913506. TSKED-SKI-----VYTYQIAFRKVDV-----SKTPLIGATF
: : : | : : | : : : : : | : : : : : | : : : : : |
gi-13621428. TAEFDSEVTFEKPQVYVYKVTEEKIDKVPQVSYDTTSYTVQVHVLWNEEQKPVATYIV
          100          110          120          130          140          150

```

```

          380          390          400          410          420          430
gi-50913506. GVDTSNKLIDIVTT-NKNGYAISTQVS-SGKYKIKELKAPKGYSLNTEYETANWVTA
| : | : | : : : : : : : : | : | : : : | : : : | : : |
gi-13621428. GYKESKVPPIQFKNSLDDSTTLTVKVKVSGTGGDRSKDFNF--GLTLKANQYYKASEKVM
          160          170          180          190          200          210

```

```

          440          450          460          470          480          490
gi-50913506. TVKTSANSKSTTYTSDKNKATDNSEQVGLKNGIFYSIDSRPTGND--VKEAYIESTKAL
| : : : : : | : : : : | : : | : : : | : : | : : |

```

FIGURE 58K

gi-19224135. YNKKGDSYNGLTETQFRAATQLAIYYFTDSTDLKTLKTYNNGKGYHGFESMDEKTLAVTK
490 500 510 520 530 540

gi-50913506.pep
/home/morama/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 60.1 E(): 7.8
>>/home/morama/gas/pili/align/gi-50913505.pep (1036 aa)
initn: 149 init1: 70 opt: 86 z-score: 60.1 expect(): 7.8
Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
(33-428:503-966)

gi-50913506. NRRETVREKILITAKKMLACLAILAVVGLGCMTRVS-ALSKDDTAQLKITNIEGGPTVTL
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNG-L
gi-50913506. YKIGEGVYNTNGDSFINFK----YAEGVSLTETGPTSQBIT-TIANGINTGKIKPFSTEN
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNRDAELLKGWSTNS
gi-50913506. VVISNGTATYNARGASVYIALITGAT-----DGRTYNPILLAAS-----YNGEGNLEVTK
gi-50913505. LLDENTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFFYFGEDGYRSG
gi-50913506. NIDS--KSNLYGQ--TSVA----KSSLPSITKKVTGTIDDVNKKTTSLSGSLVLSYSLTFE
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSPVVLKYLSE
gi-50913506. LPSY-----TKEAVNKIVY-----VSDNMSEGLTF--NFNSLIVEWKGMANITE
gi-50913505. EHYYGITDPAELEKTLNKLVEDSKLSQLGISDSLSQYVDYDQKQPDVLVTRKSKVNDETE
gi-50913506. -----DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
gi-50913505. ILYQKDQVQEAGKDIIDKVVFTPKTTTSPKGVTLTFKSDYKVDDEYTYTLTSLFNVKASDE
gi-50913506. SYKAVVNNKAIYVGEENPKAEFFYSNPNPTKGNTYDNLDKKPKD--GNGITSKEDSK----
gi-50913505. AYEKYDNEGRYSEMGSDTDYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVI
gi-50913506. IVYTYQIAFRKVDS-VSKTPLIGAIFGVYDTSNKLI-DIVTTNKNGYALSTQVSSGK-YK

FIGURE 58M

!!SEQUENCE LIST 1.0

(Peptide) FASTA of: gi-13621430.pep from: 1 to: 215 September 15, 2004 18:45

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp
Variable pamfactor used
Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences:
z-scores computed from opt scores

Table with 3 columns: z-score, obs (=), exp (*). Rows range from z-score < 20 to 100. Obs values are mostly 0, with some non-zero values (e.g., 1, 7, 21, 2, 4, 3, 4, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0). Exp values are represented by symbols like '0:', '1: *', '2: **', etc., indicating the number of search set sequences.

FIGURE 59


```

102 PCT/US2005/027239
104 0 0:
106 0 0:
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=-

```

Joining threshold: 36, opt. threshold: 24, opt. width: 16, reg.-scaled

The best scores are: init1 initn opt z-sc E(55)...

```

/home/morama/gas/pili/align/gi-13621430.pep Begin: 1 End: 215
! gi|13621430|gb|AAK33240.1| hypothet... 1338 1338 1338 233.9 1.8e-09
/home/morama/gas/pili/align/gi-19745305.pep Begin: 1 End: 193
! gi|19745305|ref|NP_606441.1| hypoth... 163 243 273 82.2 0.5
/home/morama/gas/pili/align/gi-28810261.pep Begin: 2 End: 187
! gi|28810261|dbj|BAC63199.1| hypothe... 164 239 268 81.5 0.55
/home/morama/gas/pili/align/gi-19224139.pep Begin: 2 End: 187
! gi|19224139|gb|AAL86410.1|AF447492_... 164 236 265 81.0 0.57
/home/morama/gas/pili/align/orf82.pep Begin: 30 End: 222
! TRANSLATE of: orf82.seq check: 4296... 163 235 264 81.0 0.58
/home/morama/gas/pili/align/gi-21909638.pep Begin: 2 End: 181
! gi|21909638|ref|NP_663906.1| hypoth... 164 239 261 80.5 0.62
/home/morama/gas/pili/align/gi-19745303.pep Begin: 84 End: 183
! gi|19745303|ref|NP_606439.1| hypoth... 121 121 126 61.4 6.7
/home/morama/gas/pili/align/gi-13621428.pep Begin: 6 End: 174
! gi|13621428|gb|AAK33238.1| hypothet... 58 86 122 60.9 7.2
/home/morama/gas/pili/align/gi-19224137.pep Begin: 93 End: 201
! gi|19224137|gb|AAL86408.1|AF447492_... 88 88 119 60.4 7.5
/home/morama/gas/pili/align/gi-50913503.pep Begin: 549 End: 625
! gi|50913503|ref|YP_059475.1| Fibron... 73 73 117 60.4 7.6
/home/morama/gas/pili/align/gi-19224134.pep Begin: 631 End: 697
! gi|19224134|gb|AAL86405.1|AF447492_... 73 73 115 60.1 7.8
\\End of List

```

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-13621430.pep

```

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

```

SCORES Init1: 1338 Initn: 1338 Opt: 1338 z-score: 233.9 E(-): 1.8e-09
>>/home/morama/gas/pili/align/gi-13621430.pep (215 aa)
initn: 1338 init1: 1338 opt: 1338 z-score: 233.9 expect(): 1.8e-09
Smith-Waterman score: 1338; 100.0% identity in 215 aa overlap
(1-215:1-215).

```

```

          10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
             |||
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
             |||
          10      20      30      40      50      60

          70      80      90     100     110     120
gi-13621430. EALDKESPLPNSVTTSVKNGKTSFEQLTFSEVGYHYKIHQLLGRKNSQYHYDETVYEVV
             |||
gi-13621430. EALDKESPLPNSVTTSVKNGKTSFEQLTFSEVGYHYKIHQLLGRKNSQYHYDETVYEVV
             |||

```

FIGURE 59A

(8-213-2-187) US 05/27239

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | | :: | :: | ::
gi-28810261.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTSVKGNKTSFEQLTFSEVGYHYKIHQLLGKNSQYHYDETVYEVV
      |::| : : : : :: |::|::| |::| |::| |::| : :|::| : |::|:
gi-28810261.  ESIDAMKTIEE---ITTAGSGKASESPLTFETTVGQYTYRVYQKPSQNKDYQADTTVFVDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQYSEKTPPEHPQDPTTEKEKPKQKRNIGI
      :|| |::| |::| ::::| : |::| : || : : | |::| : |::| :
gi-28810261.  VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLKPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|::: : : | : |||:
gi-28810261.  LPLAGEVKSLGILSIVLLGLLVLLYV-KKLKSRL
      160     170     180

```

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-19224139.pep
gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

```

```

SCORES  Init1: 164  Initn: 236  Opt: 265  z-score: 81.0  E(): 0.57
>>/home/morama/gas/pili/align/gi-19224139.pep (189 aa)
initn: 236 init1: 164 opt: 265 Z-score: 81.0 expect(): 0.57
Smith-Waterman score: 303; 30.6% identity in 206 aa overlap
(8-213-2-187)

```

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | | :: | :: | ::
gi-19224139.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTSVKGNKTSFEQLTFSEVGYHYKIHQLLGKNSQYHYDETVYEVV
      |::| : : : : :: |::|::| |::| |::| |::| : :|::| : |::|:
gi-19224139.  ESIDAMKTIEE---ITTAGSGKASESPLTFETTVGQYTYRVYQKPSQNKDYQADTTVFVDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQYSEKTPPEHPQDPTTEKEKPKQKRNIGI
      :|| |::| |::| ::::| : |::| : || : : | |::| : |::| :
gi-19224139.  VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKRLVLPPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|::: : : | : |||:
gi-19224139.  LPLAGEVKSLGILSIVLLGLLVLLYV-KKLKSRL
      160     170     180

```

FIGURE 59C

PCT/US05/27239

gi-13621430.pep
/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
generated symbols 1 to: 224.
GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 163 Initn: 235 Opt: 264 z-score: 81.0 E(): 0.58
>>/home/morama/gas/pili/align/orf82.pep (224 aa)
initn: 235 init1: 163 opt: 264 Z-score: 81.0 expect(): 0.58
Smith-Waterman score: 304; 30.5% identity in 213 aa overlap
(1-213:30-222)

```

gi-13621430.          10      20      30
                    MKKSILRILAI GYLLMSFCLLDSVEAENLTA
orf82.pep           LLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILTMLAFNQTVLAKDSTV
                    10      20      30      40      50      60

gi-13621430.          40      50      60      70      80      90
                    SINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVTTSVKGNKGTSTFEQLTFFS
orf82.pep           QTSISVENVLERAGDSTP-----FSVALESIDAMKTI-DEIT--IAGSGKASFSPLTFT
                    70      80      90      100     110

gi-13621430.          100     110     120     130     140     150
                    EVGQYHYKIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNKLGETEKSELIF
orf82.pep           TVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVYDDED-GTLVAKVISRRAGDEEKSAITF
                    120     130     140     150     160     170

gi-13621430.          160     170     180     190     200     210
                    KQEYSEKTPEPHQPDTEKEKPKQKRNGILPSTGEMVSYVSALGIVLVATITLYSIYKKL
orf82.pep           KPKRLVKPIPPRPQNI-----PKTP-----LPLAGEVKSLGILSIVLLGLLVLLVY-KKL
                    180     190     200     210     220

gi-13621430. KTSK
orf82.pep     KSRL

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS31
5]

SCORES Init1: 164 Initn: 239 Opt: 261 z-score: 80.5 E(): 0.62
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)
initn: 239 init1: 164 opt: 261 Z-score: 80.5 expect(): 0.62
Smith-Waterman score: 302; 31.5% identity in 200 aa overlap
(14-213:2-181)

```

gi-13621430.          10      20      30      40      50      60
                    MKKSILRILAI GYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
gi-21909638.          10      20      30      40      50      60
                    MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL

```

FIGURE 59D

PCT/US05/27239

20 30 40

```

      70      80      90      100      110      120
gi-13621430. EALDKESPLPNSVTTTSVKNGKTSFEQLTFSEVGGQYHYKIHQLLGGKNSQYHYDETVEVYEVV
      |::| : : : : : : |::|::| |::| |::| |::| |::| |::| |::| |::| |::|
gi-21909638. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
      50      60      70      80      90

```

```

      130      140      150      160      170      180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQEYSEKTPPEPHQPDTEKEKPKKRNGI
      :|| |::| |::| ::::| : : |::| : || : : | |::| : |::| : |::|
gi-21909638. VYVTFYDED-GTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPROPNI-----PKTP-----
      100      110      120      130      140

```

```

      190      200      210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKCLKTSK
      || :||: | : : |::|::| : : | : |||:
gi-21909638. LPLAGEVKSLLGILSIVLLGLLVLLYV-KKIKSRL
      150      160      170      180

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP_606439.1| hypothetical protein [Streptococcus pyogenes MGAS82 32]:

SCORES Init1: 121 Initn: 121 Opt: 126 z-score: 61.4 E(): 6.7
>>/home/morama/gas/pili/align/gi-19745303.pep (344 aa)
initn: 121 init1: 121 opt: 126 Z-score: 61.4 expect(): 6.7
Smith-Waterman score: 126; 27.0% identity in 100 aa overlap
(59-155:84-183)

```

      30      40      50      60      70      80
gi-13621430. LTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLP--NSVTTTSVKNGG-KTSF
      |::|::| : : : | | |::| : : : |
gi-19745303. LMPKADYTFKVEADDNAKGTKDKGLDIKPGVIDGLENTKTIHYGNSDKTFAKEKSVNFD
      60      70      80      90      100      110

      90      100      110      120      130      140
gi-13621430. EQLTFSEVGGQYHYKIHQLLGGKNSQYHYDETVEVYEVVIYVLYNEQSGALETNLVSNNKLGETE
      : : | || |::| : : : |::: || : | :||: |::| :||: |:::
gi-19745303. ANVKFPGVGVIYRYTVSEVNGNKAGIAYDSQQWTVVDVYVNRREDGGFPAKYIVSTEGGQSD
      120      130      140      150      160      170

      150      160      170      180      190      200
gi-13621430. KSELIFKQEYSEKTPPEPHQPDTEKEKPKKRNGILPSTGEMVSYVSALGIVLVATITLY
      |: :::|: : :
gi-19745303. KKPVLFRNFEDTTSKLVTKKVTGNTGEHQRSFSFTLLLPNECFEKGQVNVILQGGETKK
      180      190      200      210      220      230

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 58 Initn: 86 Opt: 122 z-score: 60.9 E(): 7.2
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
initn: 86 init1: 58 opt: 122 Z-score: 60.9 expect(): 7.2
Smith-Waterman score: 135; 29.1% identity in 172 aa overlap

FIGURE 59E

(8-1595617)/US05/27239

```

10      20      30      40      50
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVD-VATNKQSSDIDETFMF-
           :| | | | | :|::|::: : : : | : : | | : | : | | :
gi-13621428.  MKLRHLLLTGAALTSFAAT-TVHGETVVGAKLTVTKNLDLVNSNALIPNTDFTFKIE
           10      20      30      40      50

60      70      80      90      100
gi-13621430.  ---VIEALDK-----ESPLPN-SVTTSVKGNKTSFEQLTFSEV-----GQYHYKI-H
           | | :| : :|::: : | : | | : : : | | | | | | | | | | : :
gi-13621428.  PDTTVNEDGNKFKGVALNTPMTKVITYTNSDKGGSNTKTAEFDFSEVTFEKPGVYYYKVTET
           60      70      80      90      100      110

110     120     130     140     150     160
gi-13621430.  QLLGKNSQYHYDETVYEVVYVLYNE-QSCALETNLVSINKLGETEKSELI-FKQEYSEKTP
           : : -| | | | | : : | : | | : : | : | : | | | | : | | : : |
gi-13621428.  EKIDKVPGVSYDTSYTVQVHVLWNEEQQKPVATYIVGYKEGS--KVPIQFKNSLDSTTL
           120     130     140     150     160     170

170     180     190     200     210
gi-13621430.  EPHQPDTEKEKPKQKRRNGILPSTGEMVSYVSALGIVLVATITLYSIYKCLKTSK
gi-13621428.  TVKKKVSQGGDRSKDFNGLTLKANQYKASEKVMIEKTTKGGQAPVQTEASIDQLYHF
           180     190     200     210     220     230

```

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-19224137.pep
gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]

```

```

SCORES  Initl: 88  Initn: 88  Opt: 119  z-score: 60.4  E(): 7.5
>>/home/morama/gas/pili/align/gi-19224137.pep (342 aa)
initn: 88 initl: 88 opt: 119 Z-score: 60.4 expect(): 7.5
Smith-Waterman score: 119; 29.7% identity in 111 aa overlap
(72-176:93-201)

```

```

50      60      70      80      90
gi-13621430.  DVATNKQSSDIDETFMFVIEALDKESPLPNSVTTSVKGNK-----TSFEQLTFSEVQGY
           | : : | : : | : : | : : | : : | : : | : : | : : |
gi-19224137.  SVNPDAAATGTESNLPKPGIAVNNQDIKVSYSNTDKTSGKQVVDVFMKVTFFPSVGIY
           70      80      90      100     110     120

100     110     120     130     140     150
gi-13621430.  HYKIHQLLGKNSQYHYDETVYEVVYVLYNEQSCALETNLVSINKLGETEKSELI-FKQEY
           | : : | : : | : : | : : | : : | : : | : : | : : |
gi-19224137.  RYVVTENKGTAEQVYDDPKWLVVDVYVGNNEKGG-LEPKYLVSKKGD SATKEPLQFNNSF
           130     140     150     160     170     180

160     170     180     190     200     210
gi-13621430.  SEKTPEPHQPDTEKEKPKQKRRNGILPSTGEMVSYVSALGIVLVATITLYSIYKCLKTSK
           | | : : | : : | : : | : : | : : | : : | : : | : : |
gi-19224137.  -ETTSLKIEKEVTIGNTGDHKKAFTFTLTLQPN EY YEASSVVKIEENGQTKDVKIGEAYKF
           190     200     210     220     230     240

```

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-50913503.pep
gi|50913503|ref|YP_059475.1| Fibronectin-binding protein [Streptococcus pyogenes
MGAS10394]

```

FIGURE 59F

TYPE 3 pilus motifs
 protein F2 like fibronectin-binding protein

Length: 696-733

LPXTG

pilin motif consensus PK (X₇) K
 E box consensus ETxAPxGY

SpyM3_0104/21909640	pilin motif	155 PKEKPIIYFK
	E box	398 YTFVETAAPDGY
SPs0106/28895018	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
SpyM18_0132/19745307	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
orf84	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY

TYPE 4 pilus motifs
 protein F2 like fibronectin-binding protein

Length: 1161

LPXTG

pilin motif consensus PK (X₇₋₈) K
 E box consensus YxLxETxAPxGY

The protein is longer than the proteins belonging to TYPE 3 and has 4 possible pilin motifs and 2 E boxes

19224141	pilin motifs	215 PKGISQDIPK
		571 PKGYQOVTEK
		156 PKMSVSKY GK
		674 PKYDAKNQEVK
	E boxes	563 YDLYETKAPKGY
		940 YTFVETAAPDGY

FIGURE 60

Formation of pili structures on GBS appears to be correlated to FACS values for surface expression of GBS80 protein

WO 2006/078318

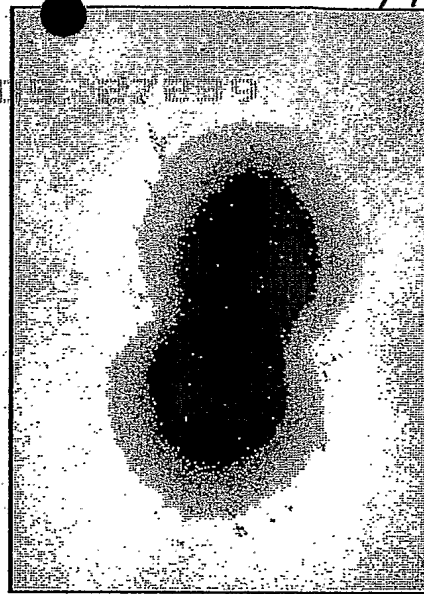
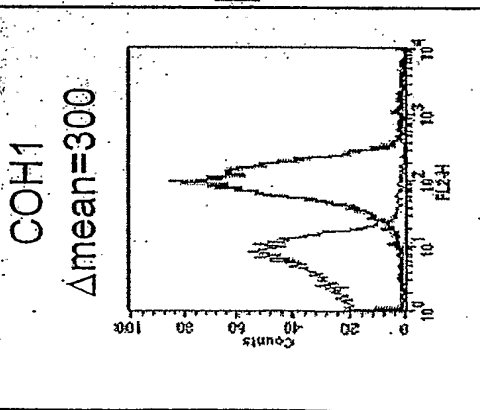
PCT/US2005/027239

211/487

PCT/US2005/027239

Immunogold Electron Microscopy

α -GBS80



α GBS80

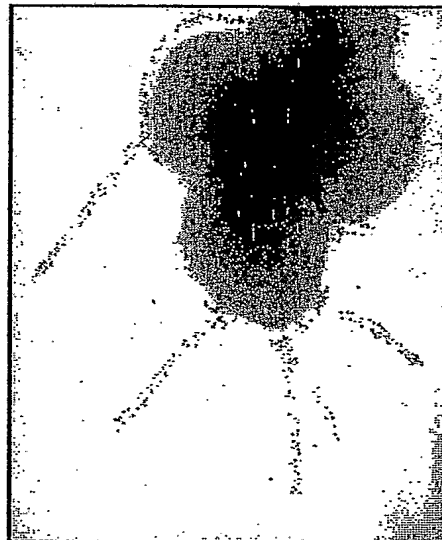
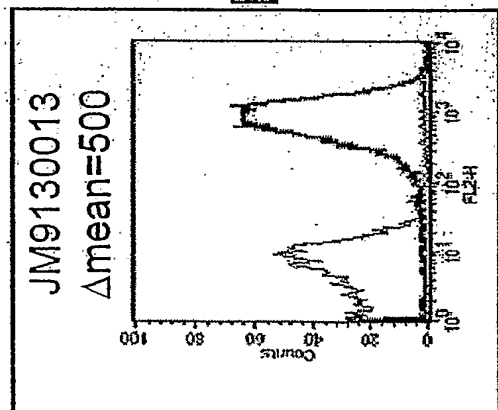


FIGURE 61

Surface exposure is capsule-dependent for GBS 322 but not for GBS 80

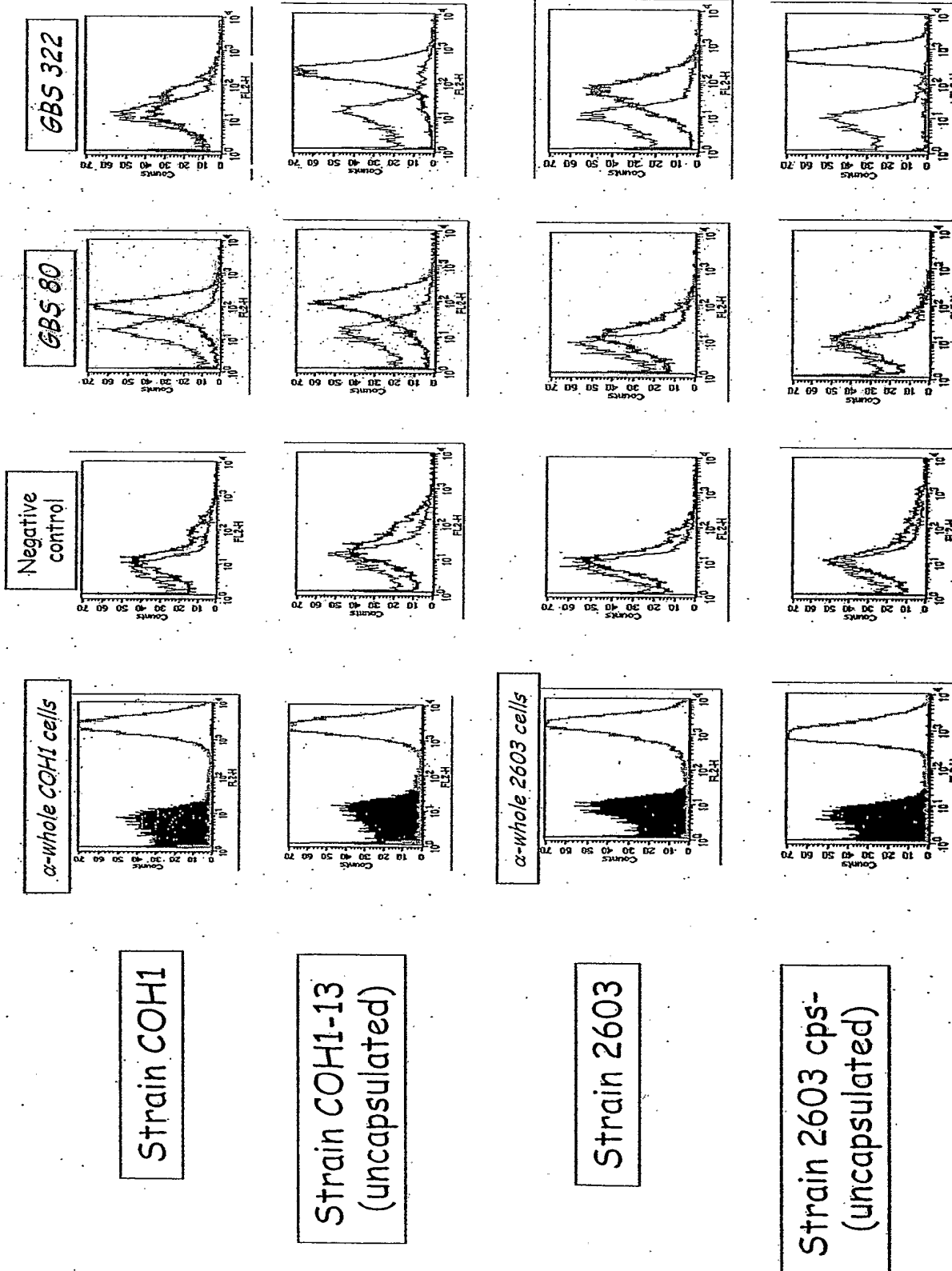
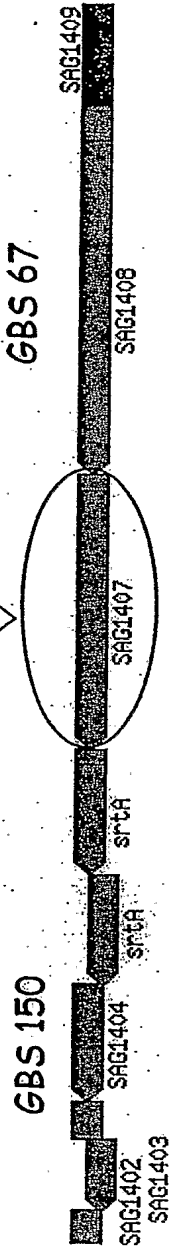


FIGURE 62

Adhesin island 2-

Operon gbs 67, 59, 150



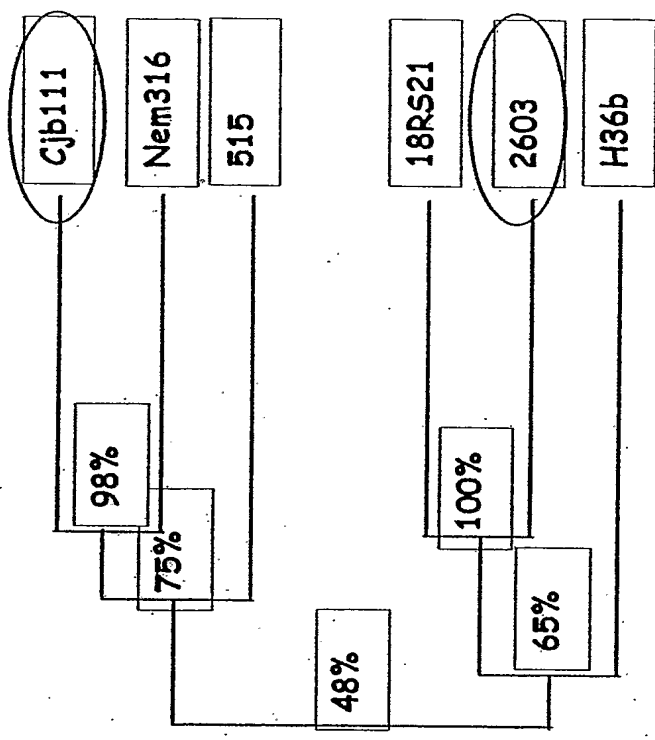
• 30 % identity with GBS 80

• By comparing GBS 59 amino acidic sequence of 2603 with that of other sequenced genomes, the following homologies are obtained.

2603	
nem316	47%
h36b	62%
515	48%
cjb111	48%
18rs21	100%
coh1	not present (Spb1)
A909	not present (Spb1)

• CGH: 1/20 GBS strains analyzed (18RS21)

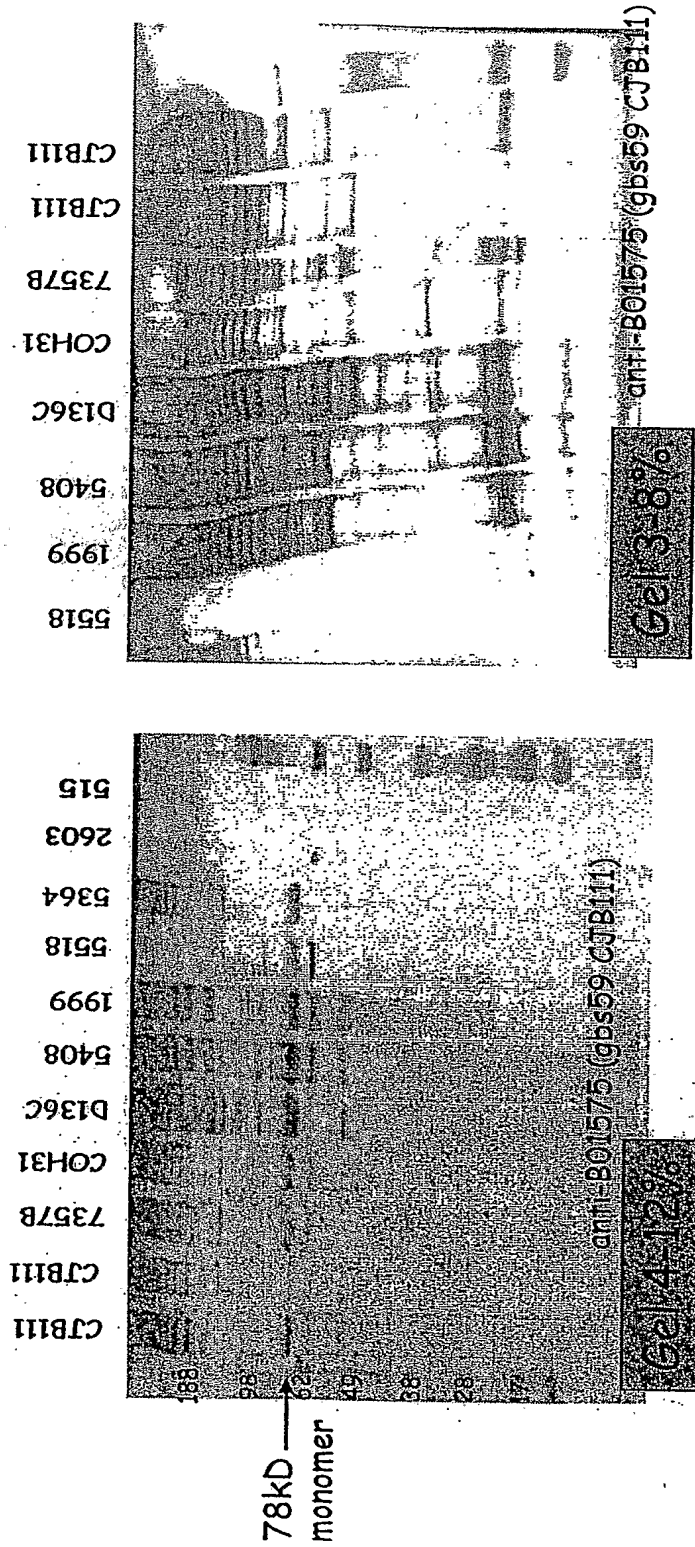
Two-by-two amino acidic sequence comparison



There seem to be two clusters suggesting the presence of two major isoforms

FIGURE 63

Western blotting with whole extracts derived from GBS strains



GBS 59 is part of a high molecular weight polymer (pilus)
 in GBS strains: CJB111, 7357b, coh31, d1363c, 5408, 1999,
 5364, 5518, 515

FIGURE 64

Western blotting with purified proteins and whole extracts derived from GBS strains

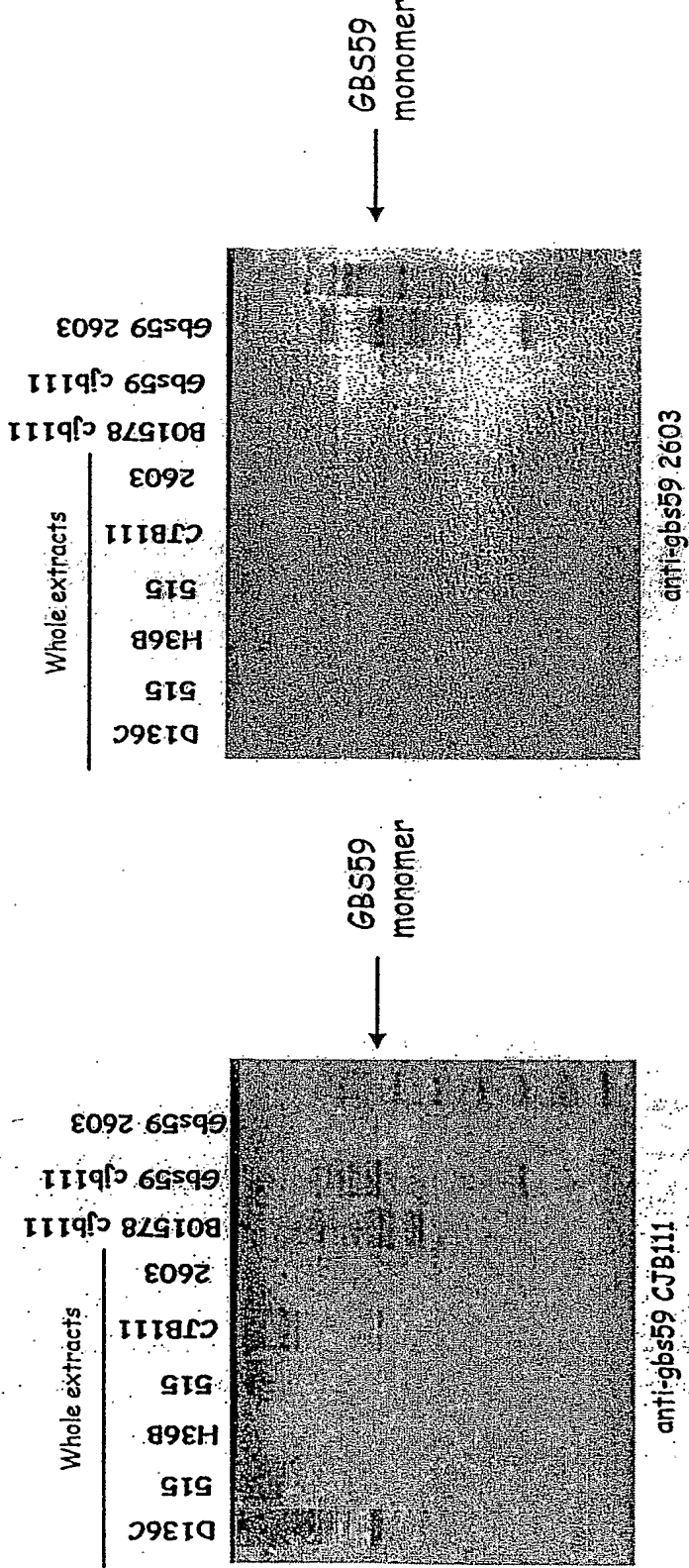
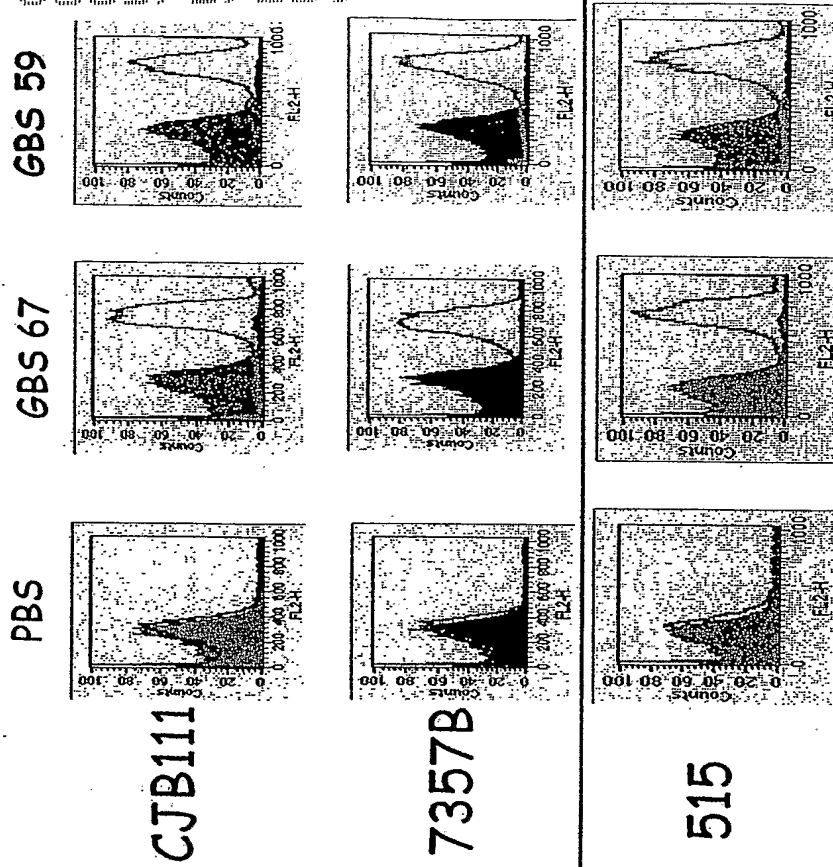


FIGURE 65

FACS analysis using mouse antiserum after immunization with BO1575 (gbs59) from CJB111 genome

GBS strains	Type	GBS 59
DK1	Ia	565
DK8		559
Davis		577
515		583
2986		443
5551		524
7357b-		596
5518		190
D136C	III	504
COH31		505
DK21	II	249
CJB111	V	493
5364		596
2110		590
1999		594
2210		636
5408		587
1169		227



Where present GBS 59 is a highly exposed on the GBS surface

FIGURE 66

Opsonophagocytosis assays: B01575 (gbs59-cjb111)

WO 2006/078318

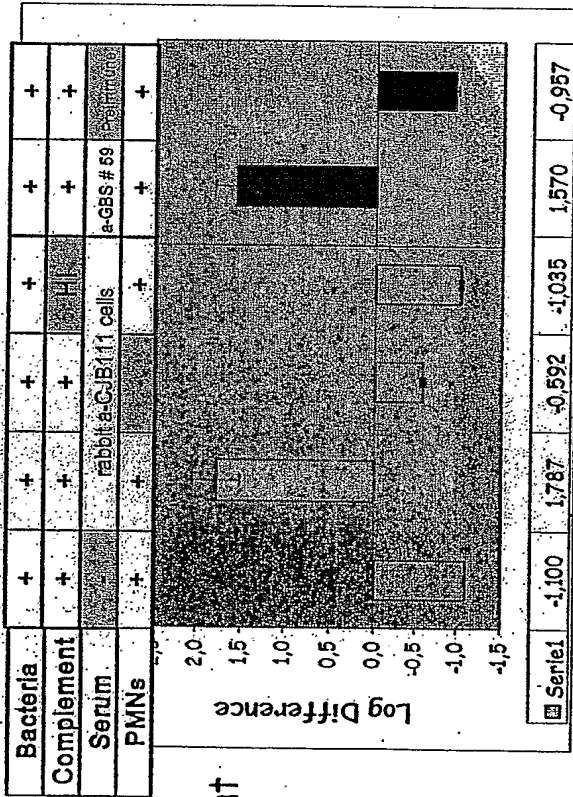
217/487

PCT/US05/27239

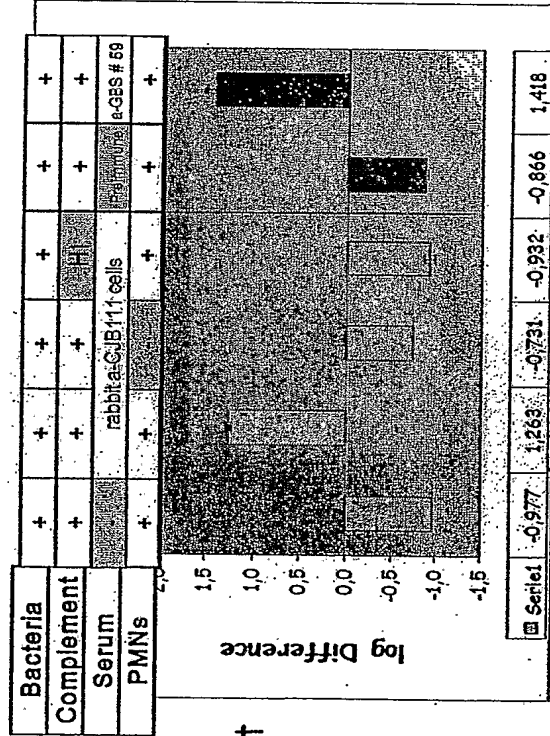
PCT/US2005/027239

- CJB111 GBS strain type Ia
- Baby rabbit complement
- Human PMNs
- Positive control: anti-type V cells (rabbit serum anti fixed type V cells)

Antibodies against B01575 (cjb111) are opsonic for cjb111 GBS strain serotype V



I experiment



II experiment

FIGURE 67

Association GBS 80-104 WB α -80, α -104 JM9130013 Total Extract

Mab α -80 A477
 Mab α -80 19G478
 Mab α -104 15H349
 Mab α -104 12A767
 Mab α -104 H2132
 Mab α -104 14B373
 α -104 POLIC.
 Mab α -80 19E177

FACS	(Δ MEAN)
GBS 80	597
GBS 104	446

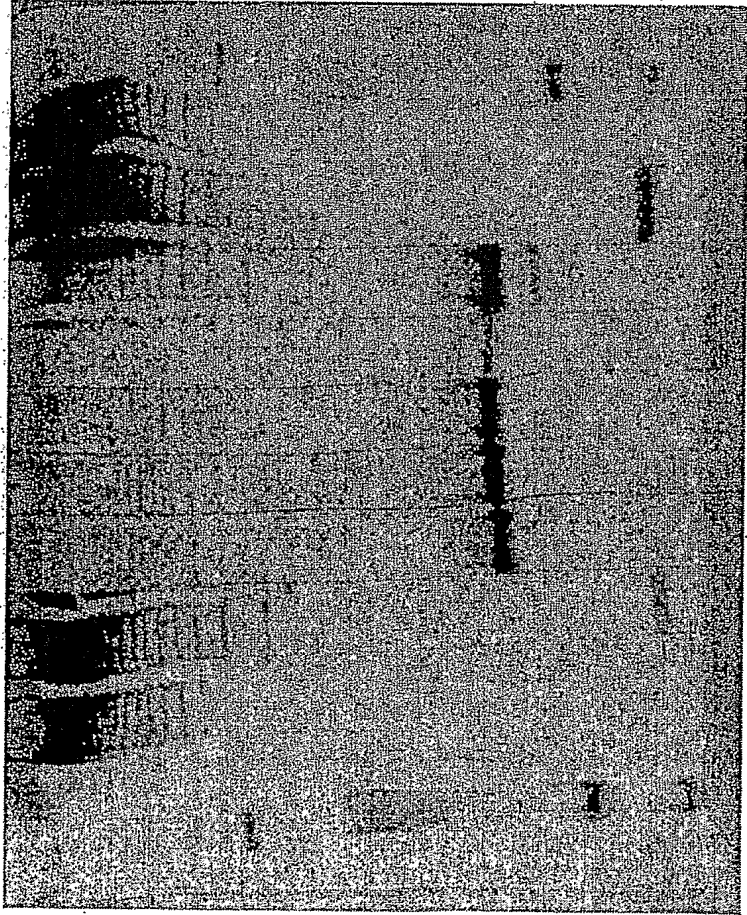
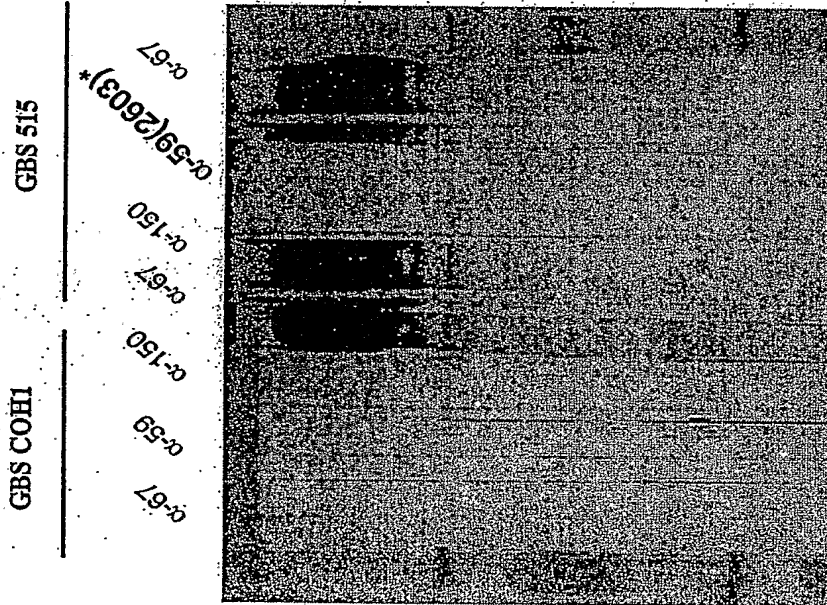


FIGURE 68

WB GBS 515 Total Extract

α -67; α -150; α -59 (2603)



Controls:
 GBS COH1 total extracts

• anti-gbs59 mouse serum after immunization with SAG1407 (GBS 59) from 2603 genome

GBS 67 and GBS 150 are parts of a high molecular weight polymer (pilus) in 515 GBS strain

FIGURE 69

Western Blotting ko GBS67
from 515 genome (clone 1.45)

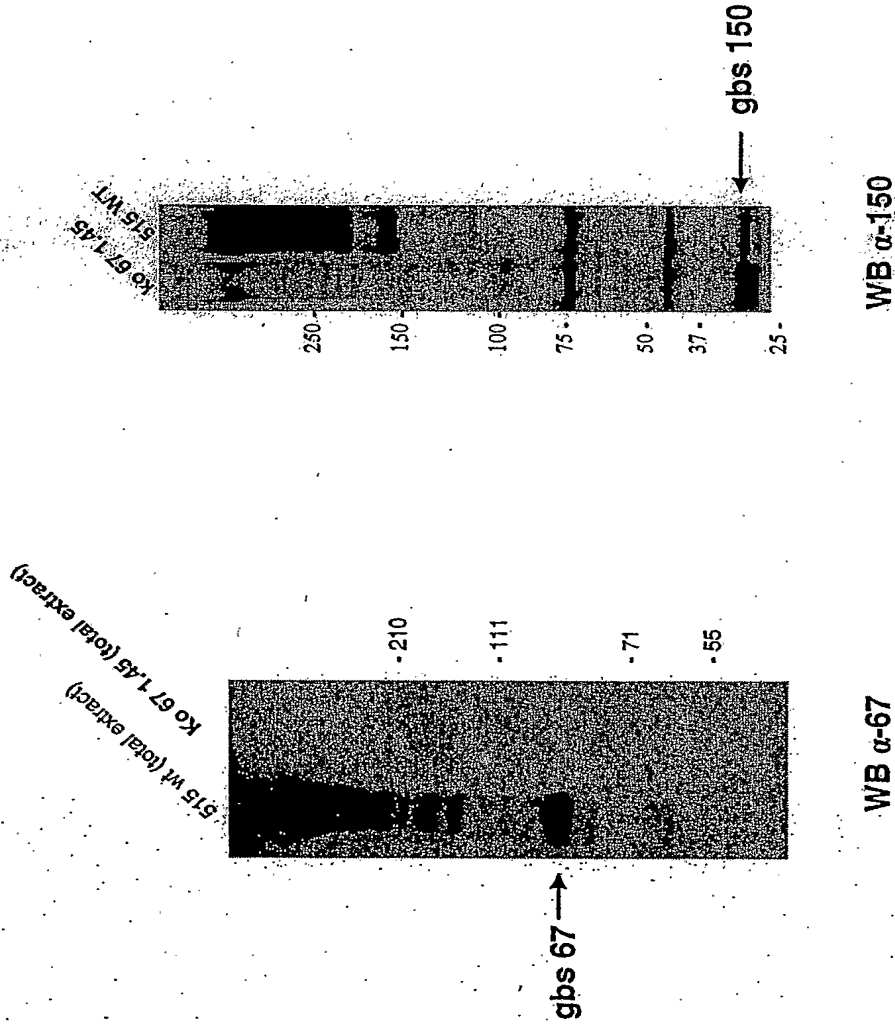
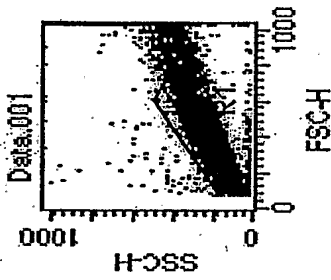
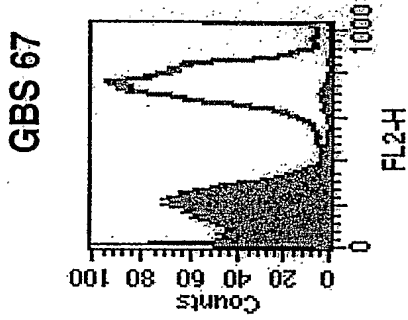


FIGURE 70

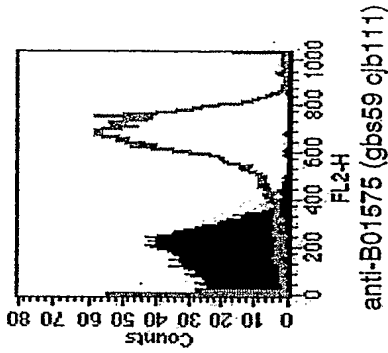
FACS GBS 515 Δ67



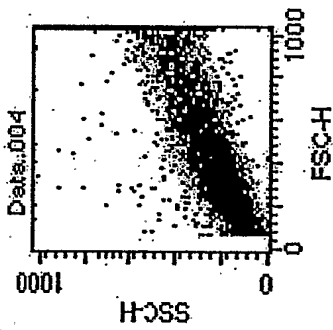
515 WT



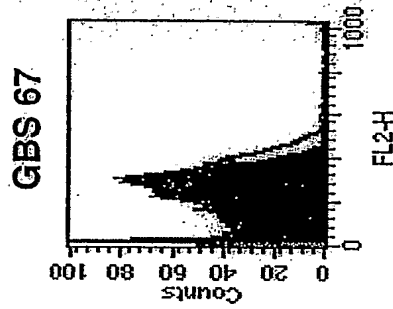
GBS 59



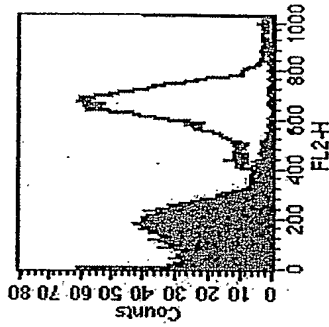
anti-B01575 (gbs59 cjb111)



515 Δ67
clone 1.45



GBS 67

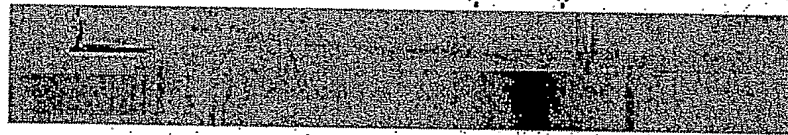


anti-B01575 (gbs59 cjb111)

FIGURE 71

Complementation of GBS 515 KO 67 with pAM401-gbs80

Ko 67 1.45
515 WT



GBS 80 forms a high molecular weight complex (pilus) in absence of GBS 67

80

WB α-80

FIGURE 72

spyM6_0159 type 1 pilus present in M6

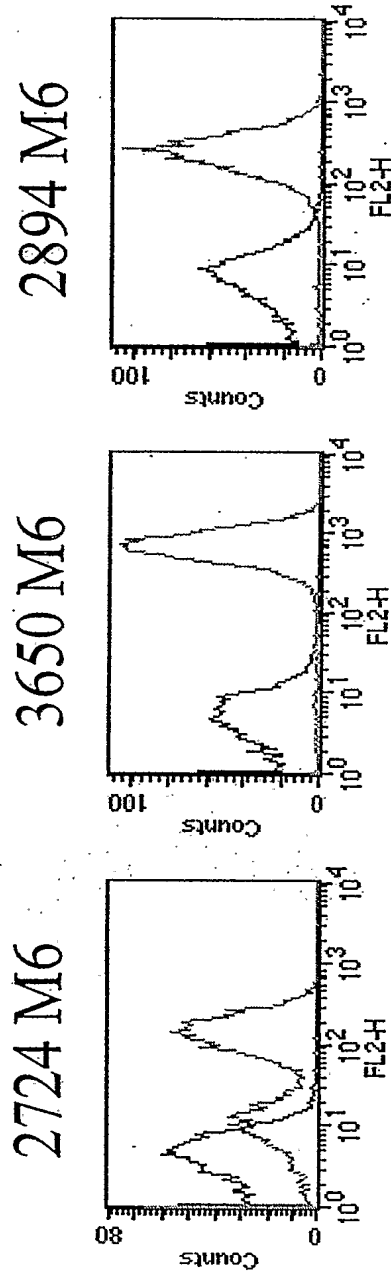


Figure 73

spyM6_0160 type 1 pilus present in M6

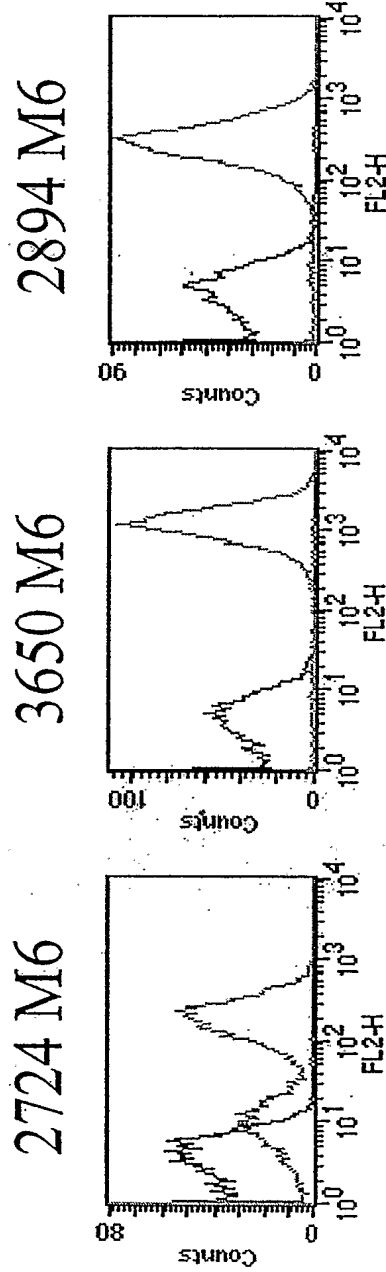


Figure 74

225/487

Gas15 type 2 pilus present in M1

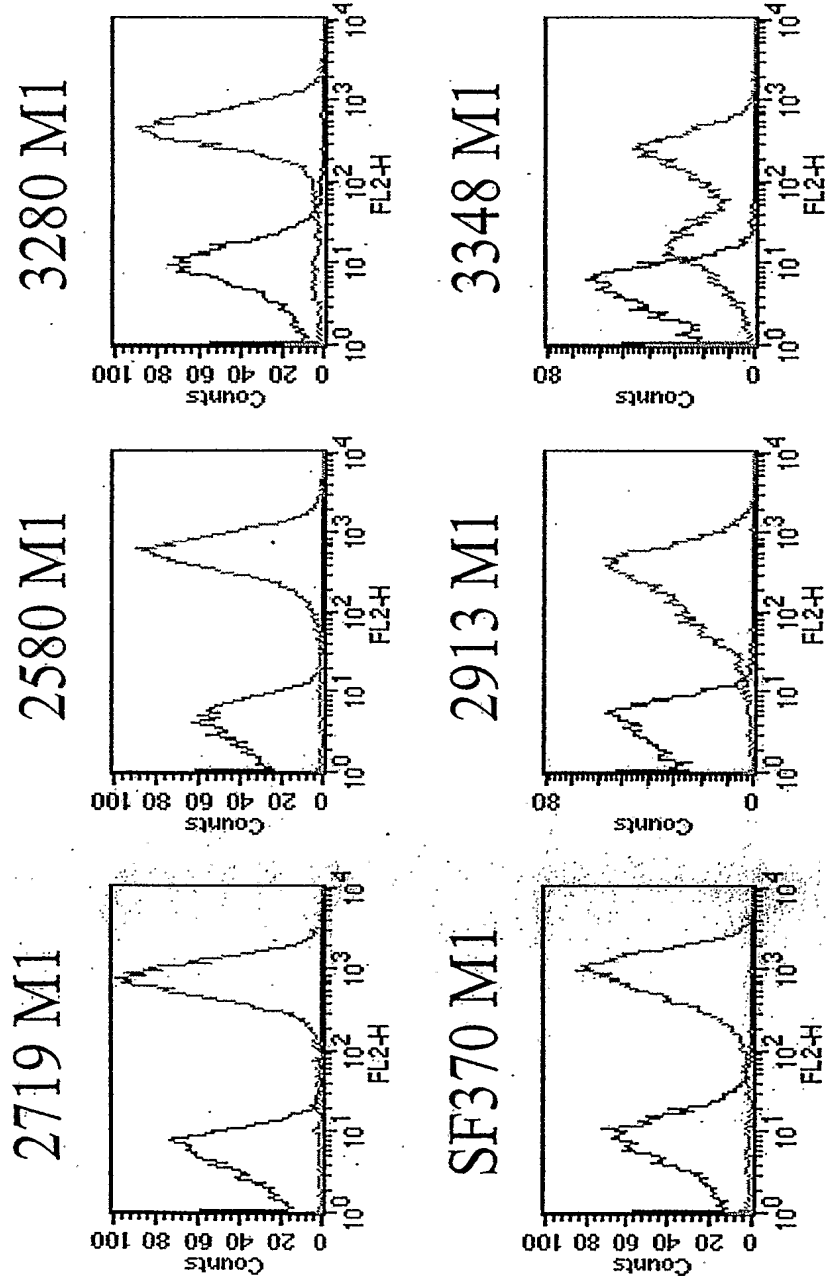


Figure 75

Gas16 type 2 pilus present in M1

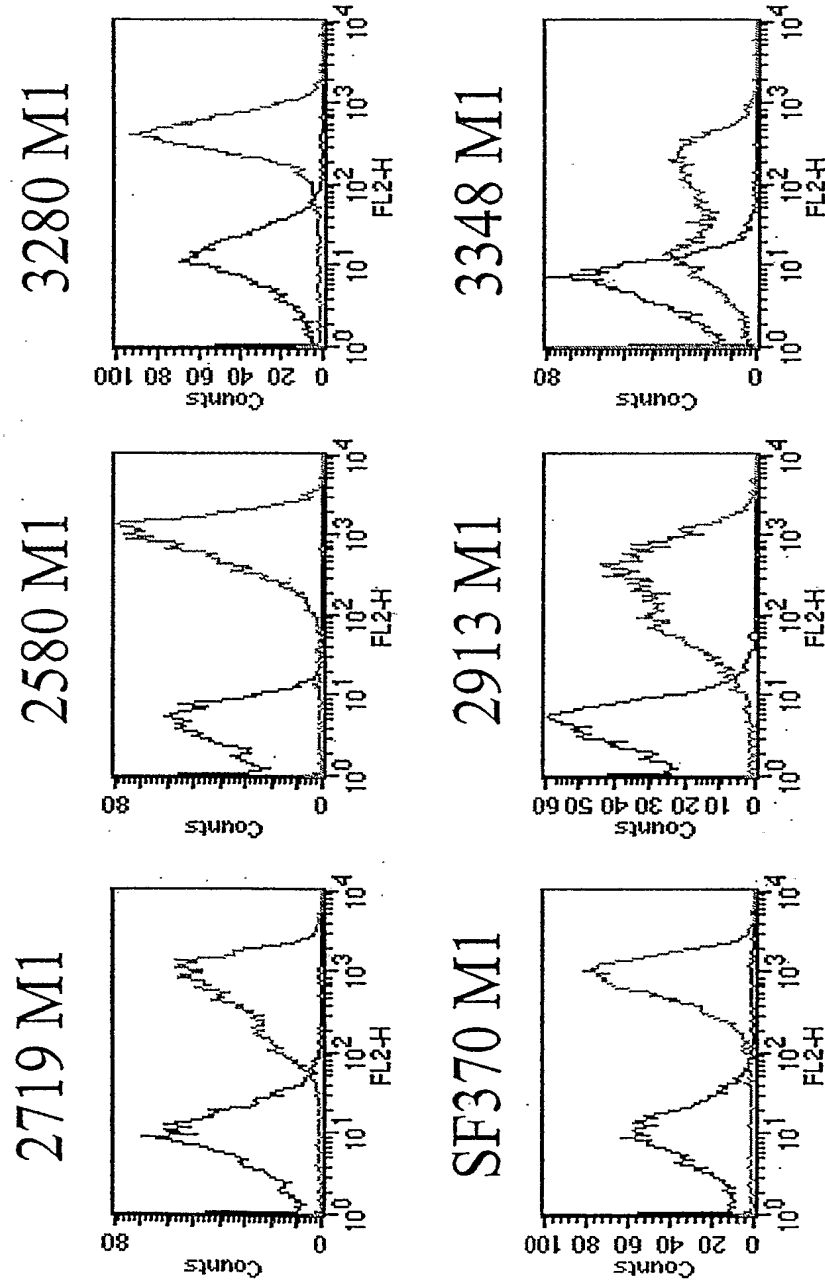


Figure 76

Gas18 serum 1 type 2 pilus present in M1

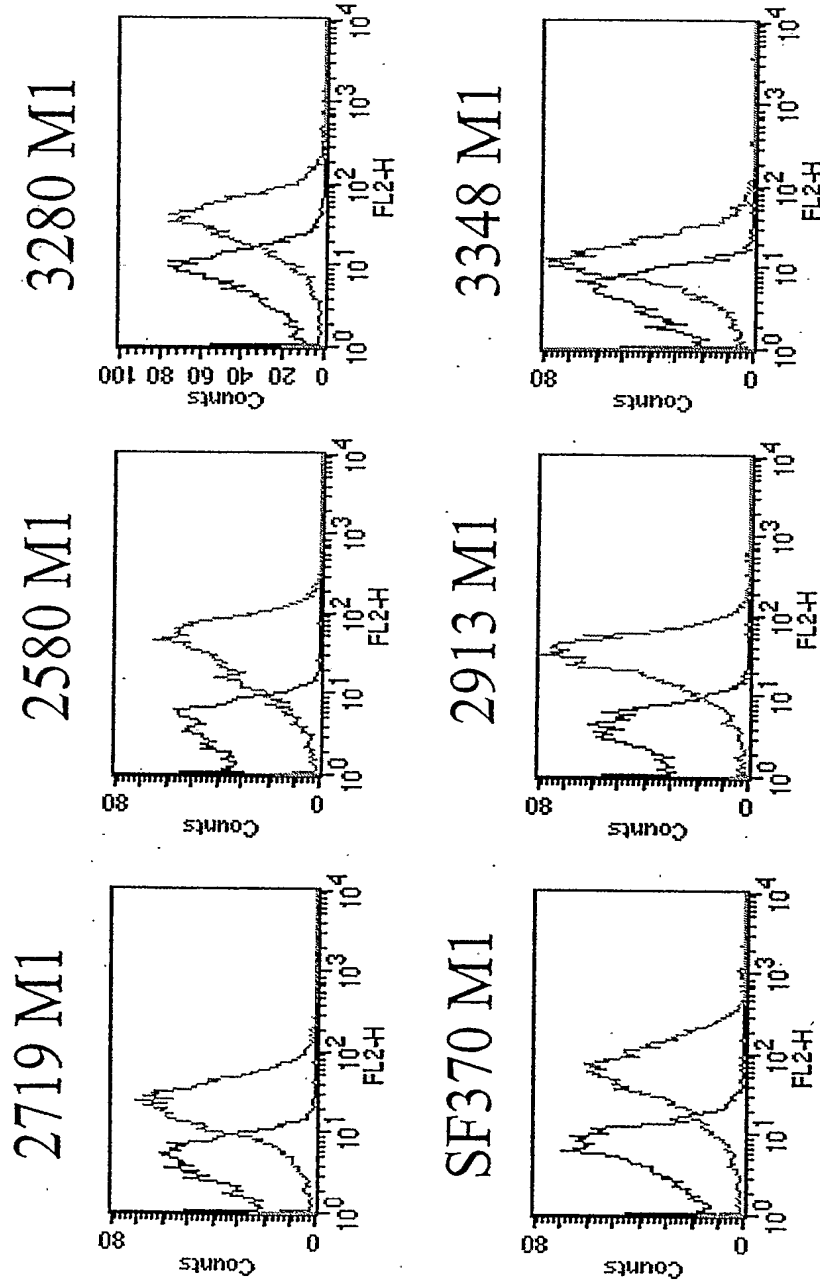


Figure 77

Gas18 serum 2 type 2 pilus present in M1

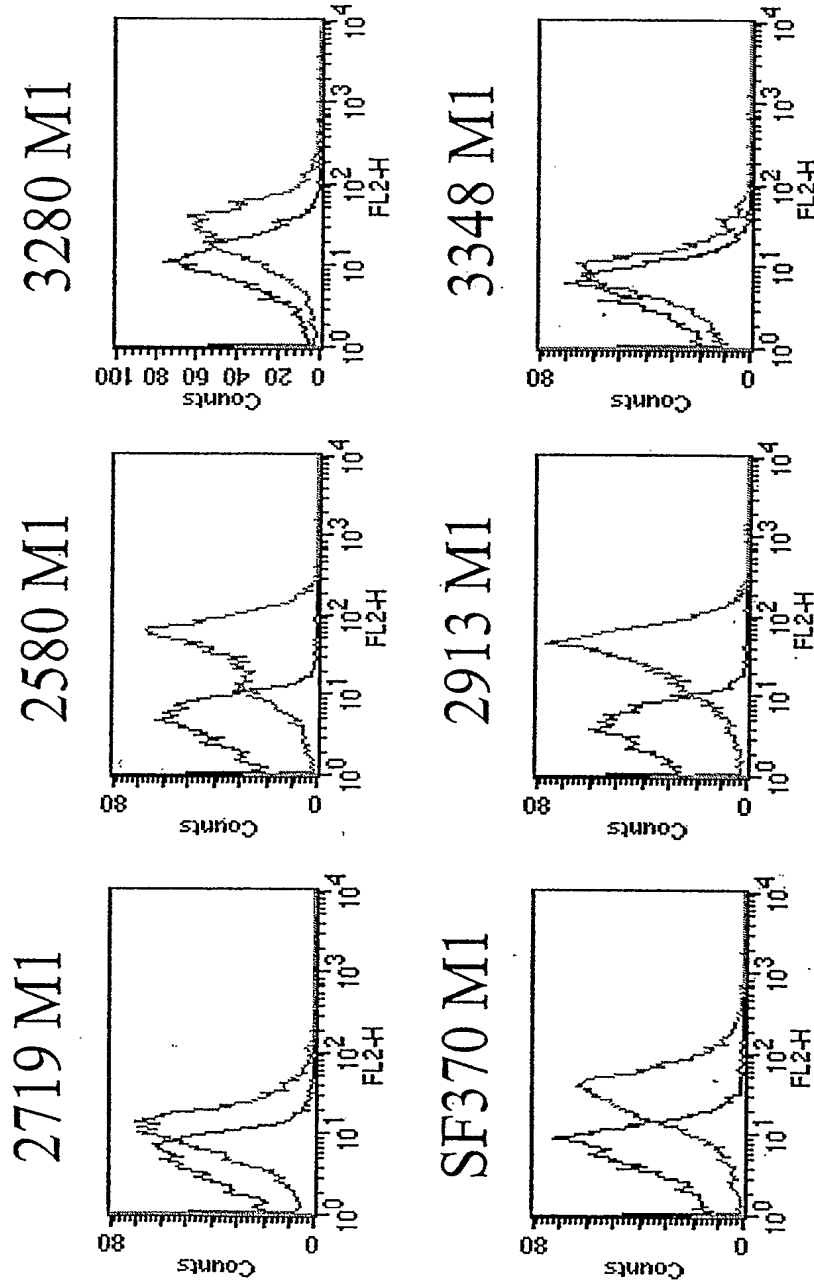


Figure 78

229/487

Gas16p2 type 2 pilus present in M1

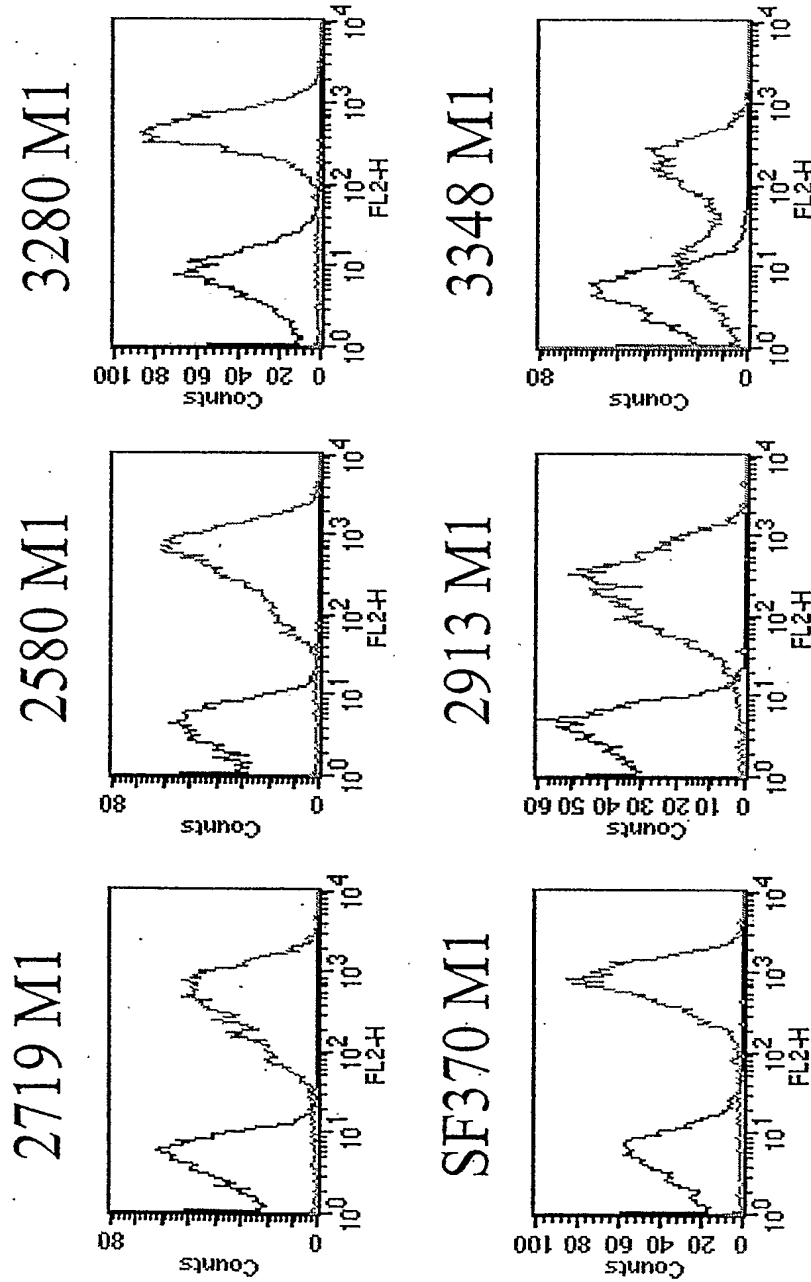
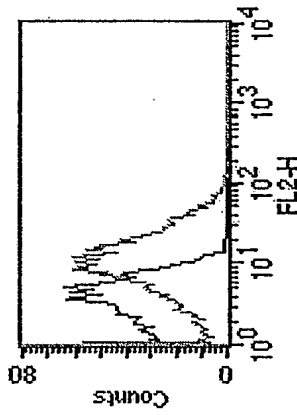


Figure 79

230/487

spyM3_0098 type 3 pilus present in M3

2721 M3



3135 M3

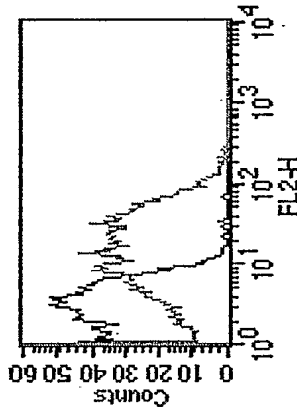
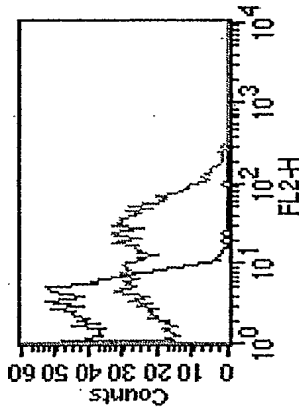


Figure 80

spyM3_0100 type 3 pilus present in M3

3135 M3



2721 M3

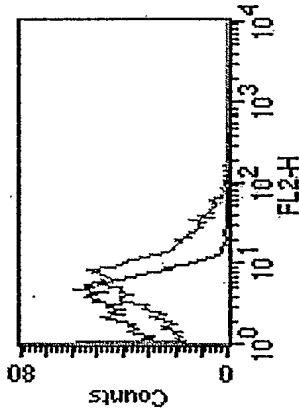
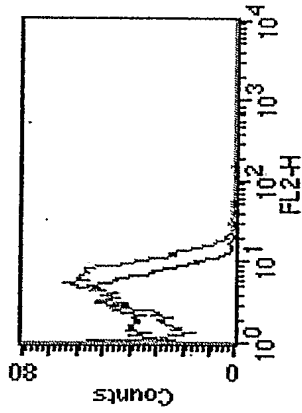


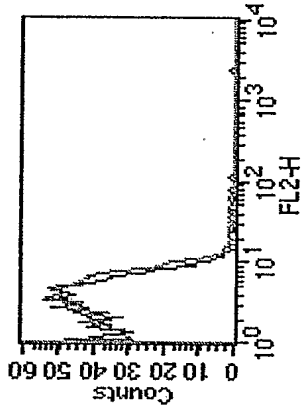
Figure 81

spyM3_0102 type 3 pilus present in M3

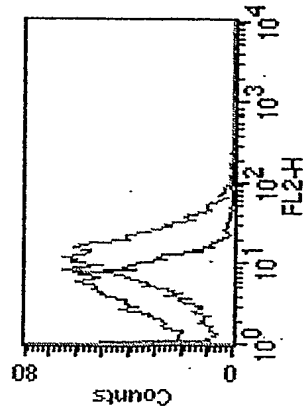
2721 M3



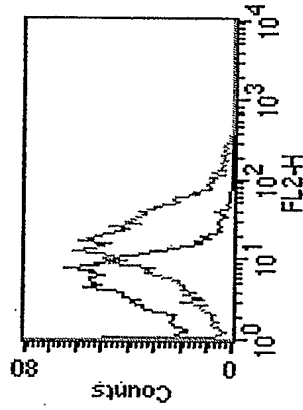
3135 M3



2724 M6



3650 M6



2894 M6

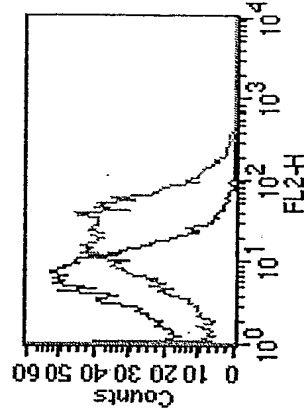
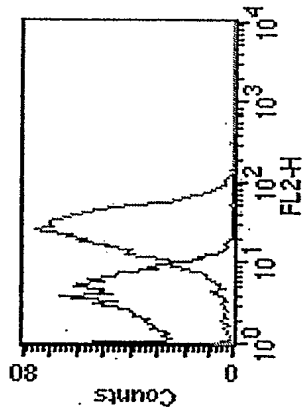


Figure 82

spyM3_0104 type 3 pilus present in M3

2721 M3



3135 M3

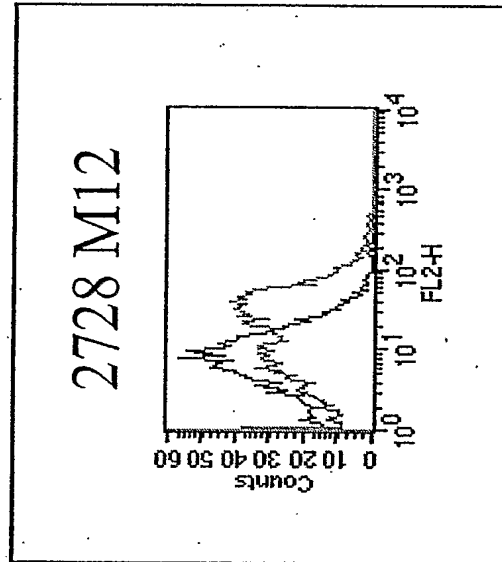
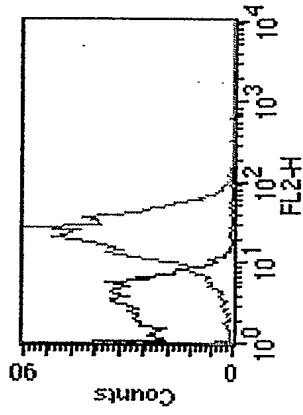
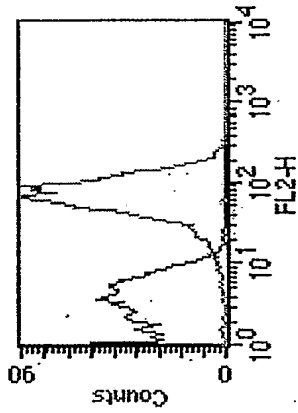


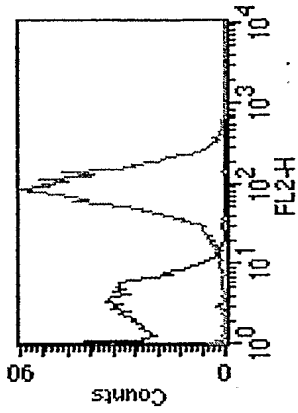
Figure 83

spyM3_0106 type 3 pilus present in M3

2721 M3



3135 M3



2728 M12

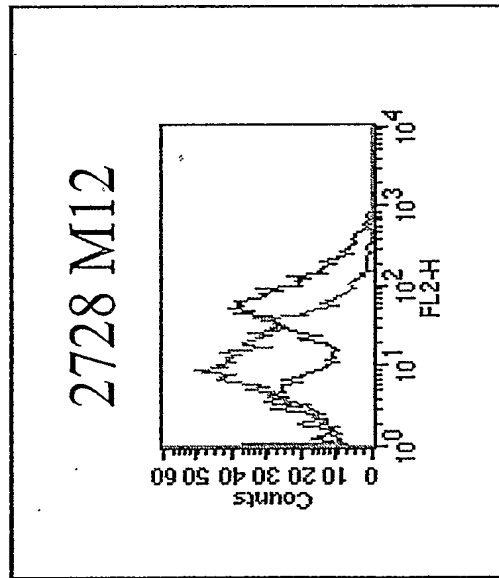
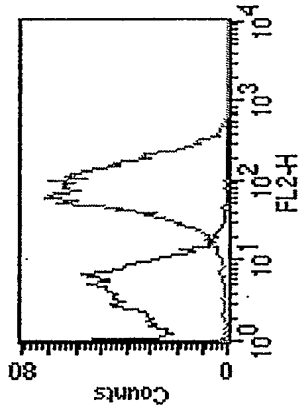


Figure 84

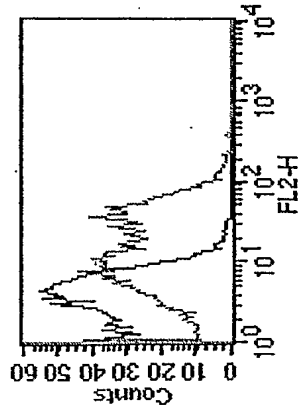
PCT/US05/27239 35/487

19224134 type 4 pilus present in M12

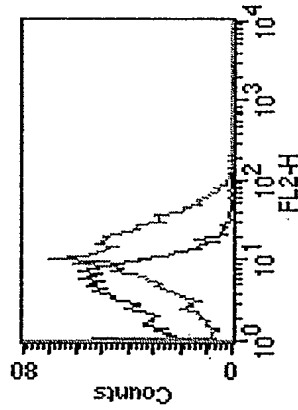
2728 M12



2724 M6



3650 M6



2894 M6

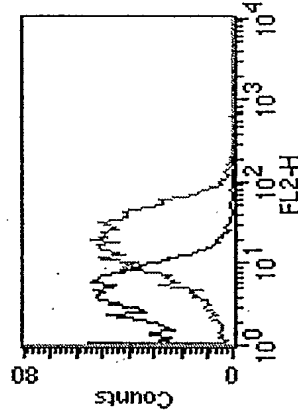


Figure 85

19224135 type 4 pilus present in M12

2728 M12

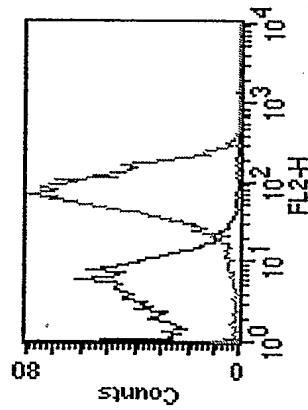


Figure 86

19224137 type 4 pilus present in M12

2728 M12

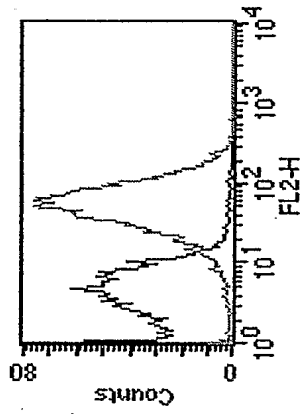


Figure 87

19224141 type 4 pilus present in M12

2728 M12

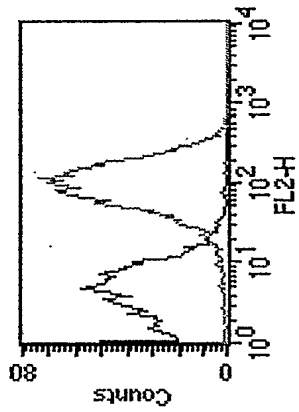
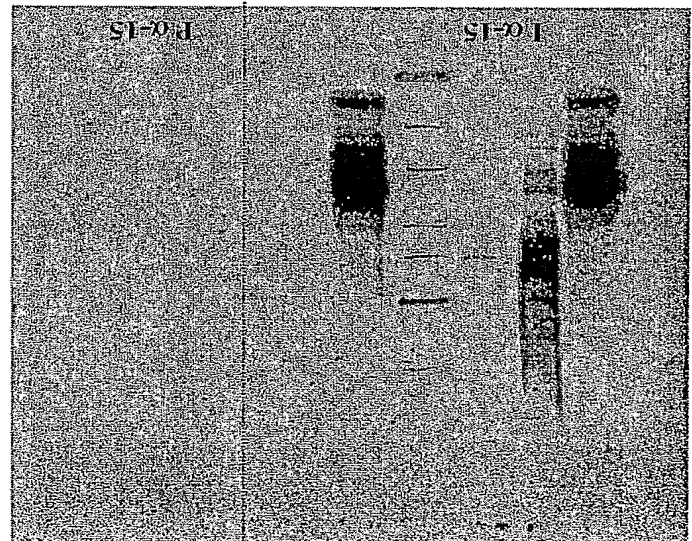


Figure 88

Figure 89

LEGEND:
 MI tot: total extract (MI)
 MI surf prot: fraction enriched in
 surface proteins
 #: Purified recombinant proteins, 30
 ng
 I α-#: immune sera against #
 P α-#: pre-immune sera

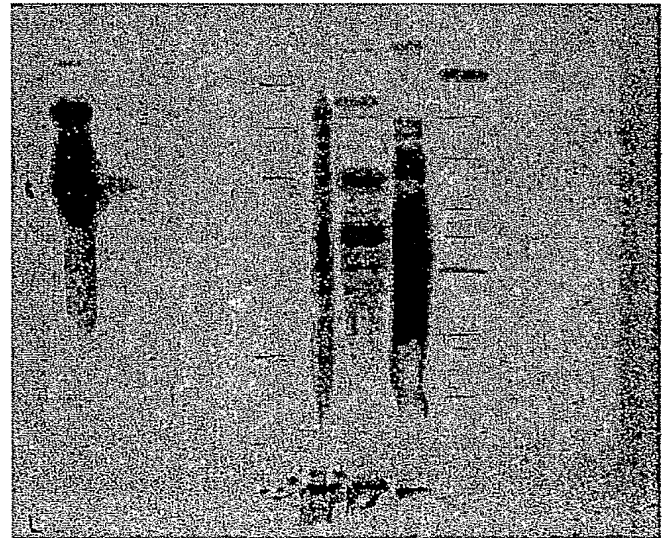


15
 MI surf prot
 MI tot
 15
 16
 18
 MI surf prot
 MI tot

240/487

Figure 90

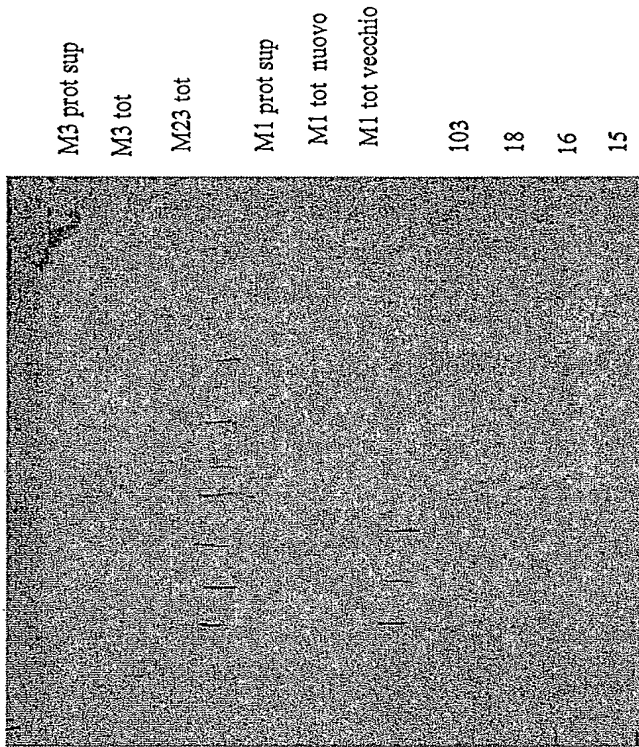
I α-15



LEGENDA:

MI tot: total extract (MI)
 MI prot. Sup.: fraction enriched in
 surface proteins
 #: Purified recombinant proteins, 30
 ng
 I α-#: immune sera against #
 P α-#: pre-immune sera

PCT/US05/27239 241/487

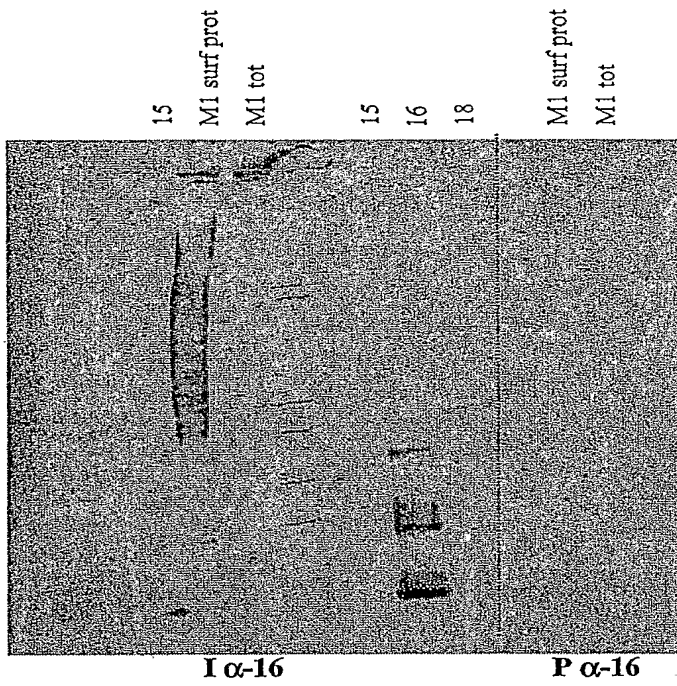


LEGENDA:
M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

P α-15

Figure 91

PCT/US05/27239 242/487



LEGEND:

M1 tot: total extract (M1)
M1 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 92

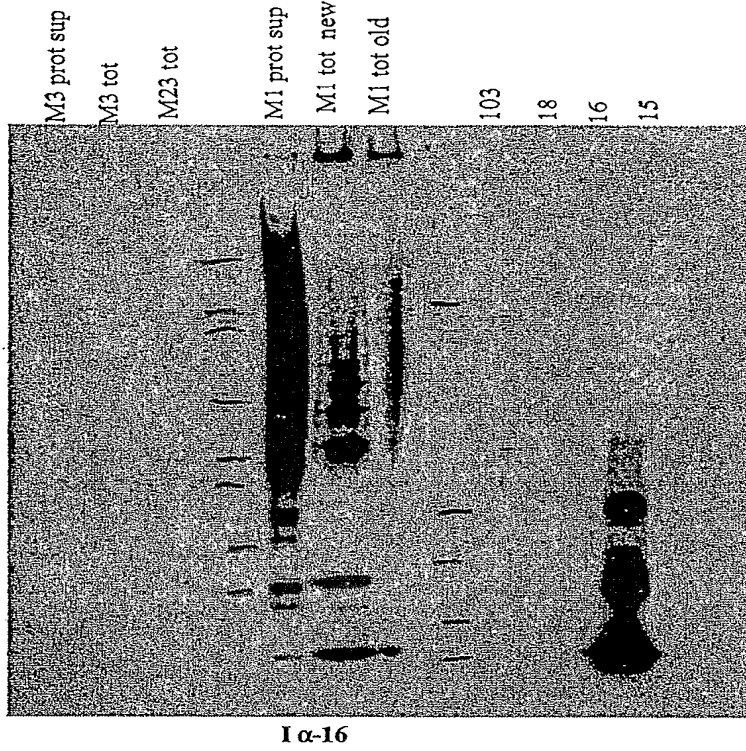
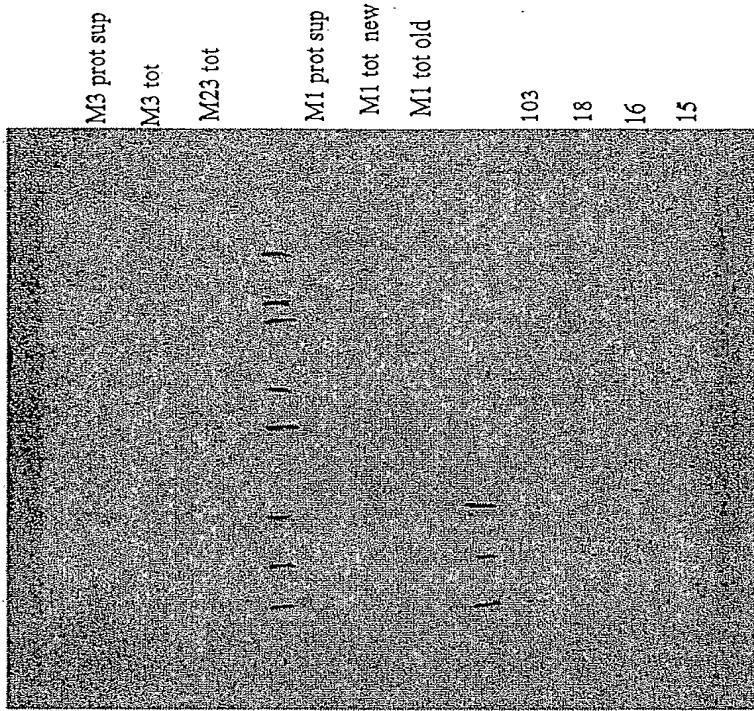


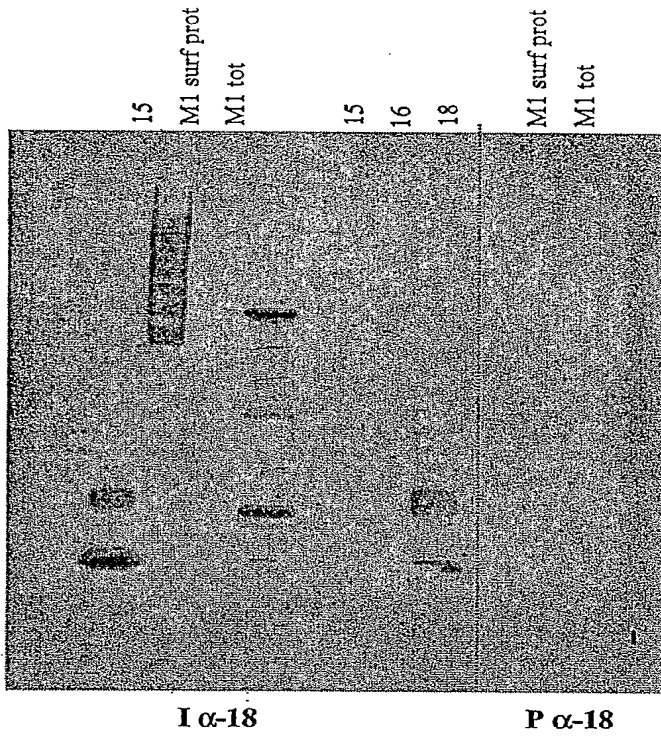
Figure 93



P α-16

LEGENDA:
M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 94

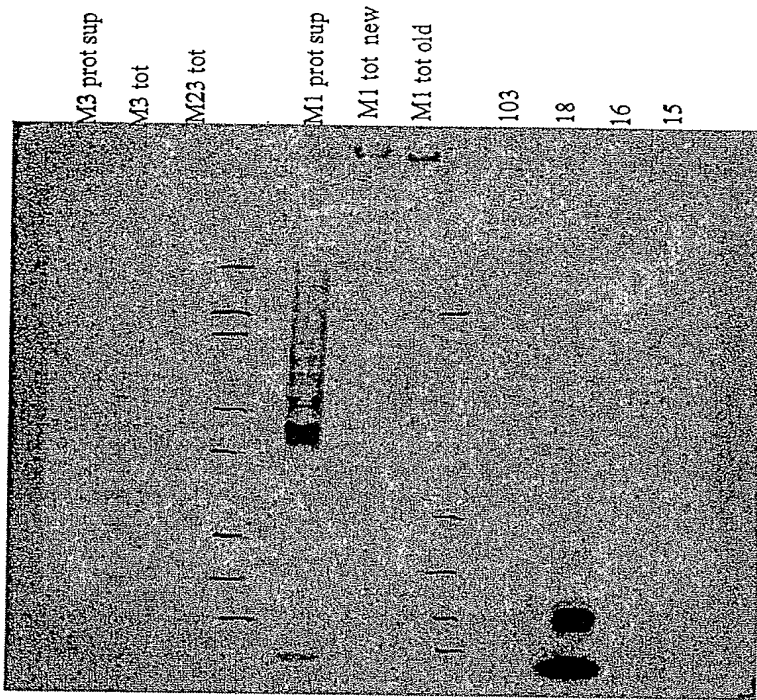


LEGEND:

M1 tot: total extract (M1)
M1 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 95

PCT/US05/27239 246/487

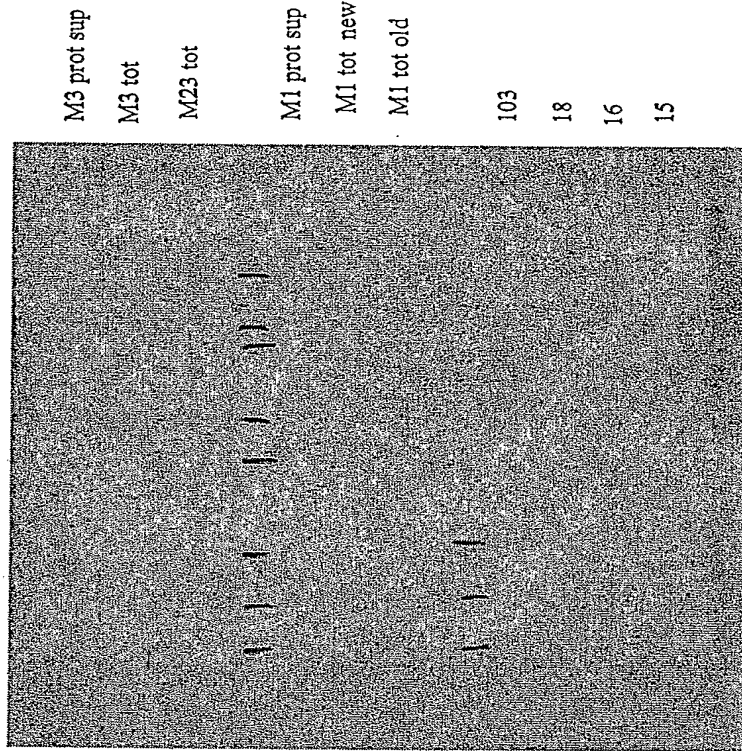


I α -18

LEGENDA:
M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α -#: immune sera against #
P α -#: pre-immune sera

Figure 96

PCT/US05/27239 247/487



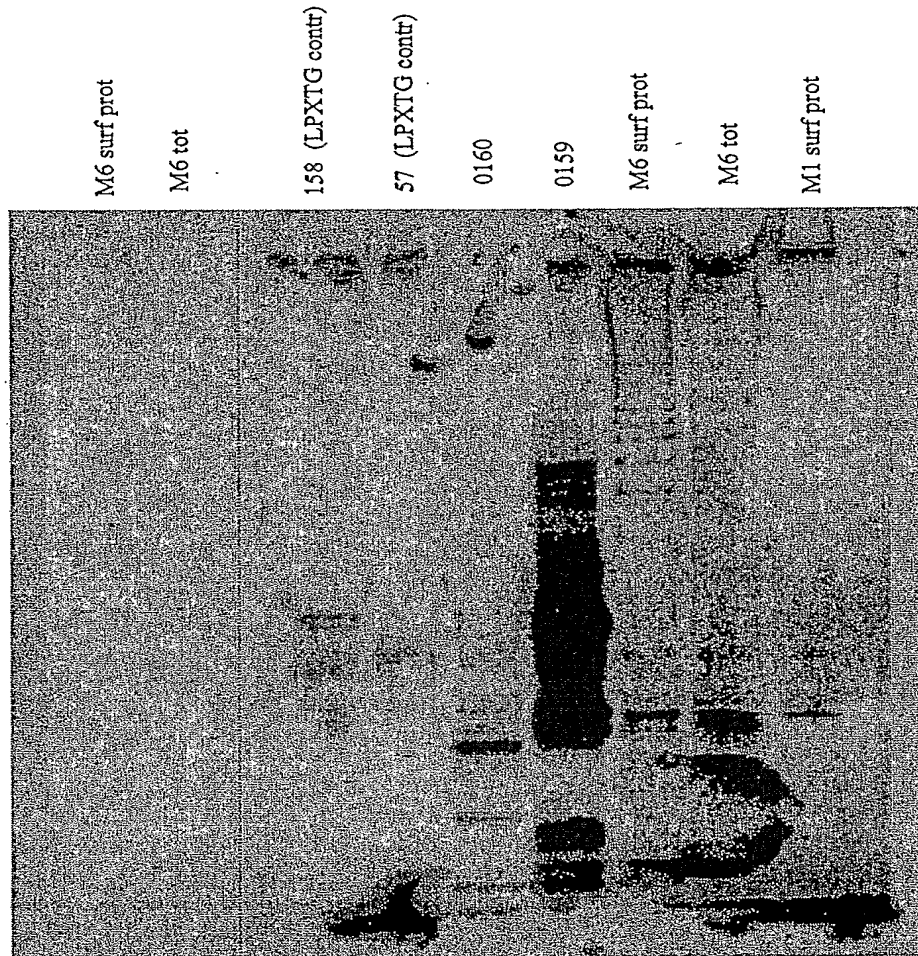
LEGENDA:
M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

P α-18

Figure 97

PCT/US05/27239 248/487

Figure 98



P α -0159

I α -0159

LEGEND:

M6 tot: total extract (M6)

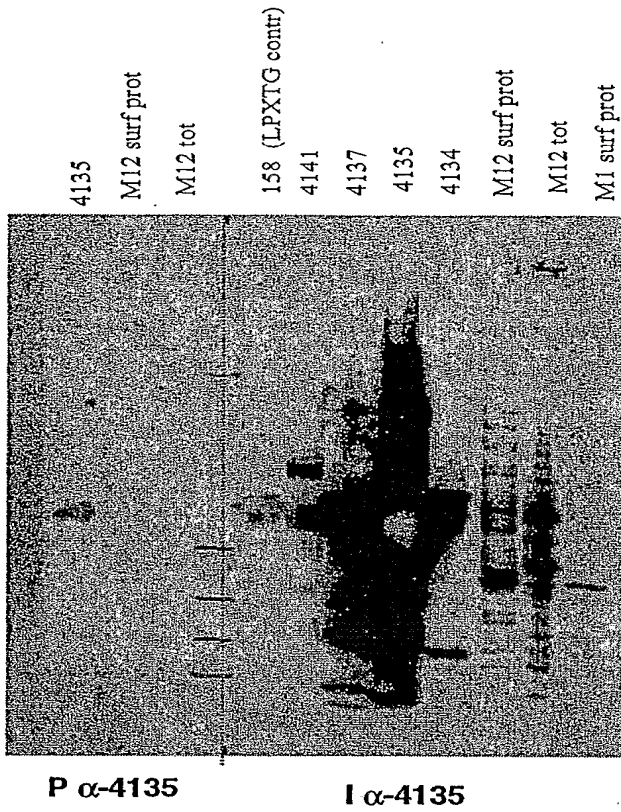
M6 surf prot.: fraction enriched in surface proteins

#: Purified recombinant proteins, 30 ng

I α -#: immune sera against #

P α -#: pre-immune sera

PCT/US05/27239 249/487

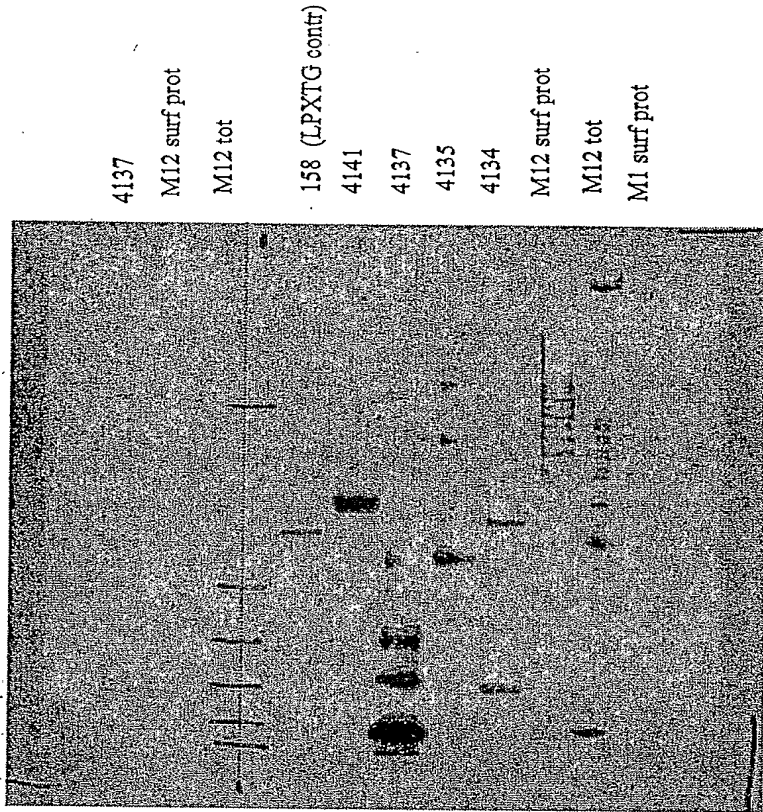


LEGEND:

M12 tot: total extract (M12)
M12 surf prot: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α -#: immune sera against #
P α -#: pre-immune sera

Figure 99

PCT/US05/27239 250/487



LEGEND:
M12 tot: total extract (M12)
M12 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

P α-4137

I α-4137

Figure 100

Figure 101A

481 TC AAAACAGTGTTTTGACAGCCCTGCGGCTAGTTCCTAGTTTCCTCTTTTCGATTTTCATTT
-----+-----+-----+-----+-----+-----+-----+-----+
AGTTTGTGCACAAAACCTCGTCGGACGGCGATCGAAGGATCAACGAGAAACTAAAAGTAA 540

541 GAGCTTTAAAATCCAGTCAGGGTAAATCCCCAAATAGCGGGACACCTCTTTCTTTCTCGCTT
-----+-----+-----+-----+-----+-----+-----+-----+
CTCGAAATTTTAGGTCAGTCCCATTTAGGGTTATCCGCCCTGTGGAGAAAGAAAGAGCGAA 600

601 AATCTTCATAGAGTTGCAGGGCTAATTTGGCTTATCTGACTAGCATCTTGTGTTTTTGG
-----+-----+-----+-----+-----+-----+-----+-----+
TTAAGAAGTATCTCAACGTCGCCGATAAACCGAATAGACTGATCGTAGAACACAAAAAACC 660

661 CAAGACTTTTTCGTTTGGTAAGAGTTGAAAAAGTCCCTCGTAGCGGATTTCAAAAATGACAA
-----+-----+-----+-----+-----+-----+-----+-----+
GTTCTGAAAAAGCAAACCATTTCTCAACTTTTCAGGACATCGCCTAAAAAGTTTTACTGTT 720

721 TTTTCCAGCTTTTCTTTGTTGATGTAGATTGAGAGCGACTTTTTCTGTAFAGAAGATCA
-----+-----+-----+-----+-----+-----+-----+-----+
AAAAAGGTCGAAAAAGAACAACTACATCTAACTCTCGCTGAAAAAAGACTATCTTCTCAGT 780

781 GCTCTTTTGTGATATCTTCTCCGACGGAGAAATCTTCCCGTAGGTTTCTCCTTGCCGA
-----+-----+-----+-----+-----+-----+-----+-----+
CGAGAAAAACTATAGAGGAGCCGTGCTCTTAGAAGGGCATCCAAAAAGAGGAACGGCT 840

841 TTGATTTACGGATCCGATTTGGATTTGACTGGAGATTGTGAATGCCACGACCCTTTCGAT
-----+-----+-----+-----+-----+-----+-----+-----+
AATAAATGCTACGCTAACCTAACTGACCCTCTCAACACTTACGGTCTCGGAAAGCTA 900

901 ACAGATCATAGCCTAGTCTACCAAAACGGTCTATTAGGGTTACTCAGGAACCTTCAAGTA
-----+-----+-----+-----+-----+-----+-----+-----+
TGTTAGTATCGGATCAGATGGTTTTTGGCAGATAATCCCAATGGAGTCTTTGAAGTTCAAT 960

253/487

Figure 101B

961 AATCAGCACCACTAATAAAACGCCCATTTGATGAAGACGCTTCTACTCTCTTTTCCCTACTC
 +-----+-----+-----+-----+-----+-----+-----+-----+
 TTAGTCGTGGTCAATTTTTCGGGTAAACTACTTCTGCAAGATGACAGAAAAAAGGATGAG
 1020
 1021 CATGAAATTTGGAAATATCCATTTGTTGAGAAAACTCTCAGCCCTGTTTCAGGTAGAAATCA
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GTACTTTAAACCTTTATAGGTAACAACACTCTTTTAGGAGTCGGACAAGTCGATCTTAGT
 1080
 1081 CTGTCAAAACCACTGTGGTTTTTGATAATCACTCGCCATTTTAGCTAAGAAATTTGTGTAAAG
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GACAGTTTGGTACACCAAAAACTATTTAGTGAGCGGTAATAATCGATCTTTAAACAACATTC
 1140
 1141 AAACGCCCTGCGGAAGCAGTTAGATGGAGTCTTCCAGATAATCTTTTGAATGAGGGCAG
 +-----+-----+-----+-----+-----+-----+-----+-----+
 TTTCCGGACGCCCTCGTCAATCTACCTCAAGAAAGGCTATAGAAAAAACTTACTCCGCTC
 1200
 1201 CAATTTGACCGCTGACTTGATACCGAGTTATTTCTGTCTCACATCCAAATAGGCTTCTGT
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GTTAAAACTGGCGACTGAACATATGGCTCAATAAATAAGACAGTGTAGGTTTATFCCGAAGCA
 1260
 1261 CAATGCTCATGGGTTCAATCAAAATCTGTATAGCGCTTAAAAATAAGACTCGAATCCGGAGTC
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GTTACGAGTACCCCAAGTTAGTTTAGACATATCGGAAATTTTATCGAGCTTAGGCCCTCAG
 1320
 1321 CCACAGACTTGTATTTCTCATAAATCCCTGAGATAAAGACAGCCCTGGGGACAACCGTTTCAAT
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GGTGCTGAACATAAAGAGTATTAAGGACTCTATTTCTGTGCGGACCCCTGTGTGCAAGTA
 1380
 1381 AAGCTTCCTTGGAACTCATGGCAGAAATGGACACCAAAAGCTCTTTCCTCATTAACACTACAG
 +-----+-----+-----+-----+-----+-----+-----+-----+
 TTCGAAGGAACCTTGATACCGTCTTACCTGTGGTTTTTCGAGAACGGAGTATGATGATGCC
 1440

PCT/US05/27239 254/487

Figure 101C

1441 TAGAACGACTCCCGTCCACCCTGTTTCCGAGGGTCCCTTCCAAATAATGACAGGTTTTC
-----+-----+-----+-----+-----+-----+-----+
ATCTTTGCTGAGGGCAGGTGGACAAACGGCTCCAGCGAAGGTTATTACTGTCCAAAAG 1500

1501 CTCTGAGTTTAGGATATCCCTGATTTCCACTGCAGCAAAAAGGCATCCATGTCATAT
-----+-----+-----+-----+-----+-----+-----+
GAGACTCAAATCCTAATAGGGACTAAGGTGACGTCGTTTTTCCGTAGGTACAGTTATA 1560

1561 GGATGATTTTCTTGACAAATCATTTAACAAAGGAAAAATCAACATGCCCTAGCACCTTTT
-----+-----+-----+-----+-----+-----+-----+
CCTACTAAAAGAACTGTTTAGTAAATTTGTTTCCCTTTTAGTTGTACGGATCGTGGAAAA 1620

1621 TATACTCTTCGAAAAATCTCTCAAAACCAGTCCAGCTTCCATCTGCAACCTCAAAACAGTA
-----+-----+-----+-----+-----+-----+-----+
ATATGAGAAGCTTTTAGAGAAGTTTGGTCCAGTCGGAAGTAGACCTTGGAGTTTGTCAAT 1680

1681 TTTTGAGCTGACTCGTCACTTCTATTACAAACCTCAAGCAGTCTTTGACCAGCCTGC
-----+-----+-----+-----+-----+-----+-----+
AAAACTCGACTGAAGCAGTCAAGATAAATGTTGGAGTTTCGTCACGAAACCTCGTCGGGAGG 1740

1741 GGCTAGTTTCCCTAGTTTGGCTTTTCGATTTCCATTTGAGTGTAACTGCTTATTTTCTTTTAT
-----+-----+-----+-----+-----+-----+-----+
CCGATCAAAGGATCAAACGAAAAGCTAAAGGTAACACTCACATTCAGCGAATAAAGAAAATA 1800

1801 TATAACCTTTTTCGAAAAAAGAAAAAGGACTTTATTTTTCAAAAATATAATACA
-----+-----+-----+-----+-----+-----+-----+
ATATGGGAAAAAAGACTTTTTCCTTTTTCCTGAAATAAAAAAAGTTTTTATATATANGT 1860

1861 GTTTGAATAAATAATAGACTGTTTGTAGAAAAAGTAAGTGTAAAAATAGGAATTTTTCACCT
-----+-----+-----+-----+-----+-----+-----+
CAAACCTTATTTTATATCTGACAAAATCTTTTCTTTTTCACATTTTATCTCCTTAAAAAGTGA 1920

Figure 101D

1921 TGTTGAAATCGGTTACTTTATGGTATACCTTGTCTCANGAATGTAAACAGATGACTGTTACT + 1980
 ACAAATTTAGCCAAATGAATACCAATATGAACAGAGTACTTACATTTGTTACTGACAAATGA
 AGAAAAAGAGGACATTAATATGGTTGTTAAGACAGCTTGTGAAGCACACAAGATATTTTTC
 1981 +-----+-----+-----+-----+-----+ 2040
 TCTTTTTCCTGTAATTATACCAACAATCTGTCAACAACACTTGGTGTCTATATAAAAAAC
 M V V K T V V E A Q D I F D - this orf is the homologue of sp0459, a formate
 acetyltransferase (pfl). It is out of the
 pilus locus
 2041 ACAAAAGCTTGGGAAGGCTTCAAAAGCGTAGATTTGAAAGAAAAAAGCAAGTATCACGGCT + 2100
 TGTTCGAAACCCCTTCCGAAAGTTTCCGGATCTAACCTTTCTTTTTCGTTACATAGTGCGA
 K A W E G F K G V D W K E K A S V S R F -
 2101 TTGTACAAGCTAACTACACACCTTATGATGGAGACGAAAGCTTCTTCAGGACCAACAG + 2160
 AACATGTTGATGATGTTGGAAATACTACCTCTGCTTTCGAAAGAACGCTCTGGTTGTC
 V Q A N Y T P Y D G D E S F L A G P T E -
 2161 ACCGTTCACTCACATCAGAAAATTTGAGAAACTAAAGCACACTACGAAAGAACTC + 2220
 TCGCAAGTGAAGTGTAGTCTTTTAAACATCTTCTTTGATTTCCGTTGATGCTTCTTTGAG
 R S L H I K K I V E E T K A H Y E E T R -
 2221 GTTCCCAATGGACACTCGTCCAACATCTATCGCTGATATCCCTGCTGGATTTATCGACA + 2280
 CAAAGGTTACCTGTGACAGGTTGTAGATAGCGACTATAGGACCGACCTAAATAGCTGT
 F P M D T R P T S I A D I P A G F I D K -
 2281 AAGAAAATGAAGTTATCTTTGGTATCCAAAATGATGAACCTCTCAAATTTGAACCTCATGC + 2340
 TTCTTTTACTTCAATAGAAAACCATAGGTTTACTACTTGTAGAAAGTTTAACTTGAAGTACG

Figure 101E

C E N E V I F G I Q N D E L F K L N F M P -
 2341 CAAAAGGTGGTATCCGTATGGCTGAAACTACTTTAAAGAAAATGGATACGAACCCAGACC + 2400
 GTTTTCCACCATAGGCATACCGACTTTGATGAAATTTCTTTTACCTATGCTTGGTCTGG
 C K G G I R M A E T T L K E N G Y E P D P -
 2461 CAGCTGTTCCAGCAAATCTCAATAATATGTAACAACAGTTAACGACGGTATTTCCGGTG + 2460
 GTCGACAAGTCTTTAGAAGTGATTTATACATTTGTTGTCAATGCTCAATTGCTGCCATRAAAGGCAC
 C A V H E I F T K Y V T T V N D G I F R A -
 2461 CCTACACTCAAAATATTCGTCGGCTCGTCAAGCACACACTGTAACCTGGTCTTCCAGATG + 2520
 GGATGTGAAGTTTATAACGACGCCGACAGTACCTGTGTGACATGACCAGAGAAGGCTTAC
 C Y T S N I R R A R H A H T V T G L P D A -
 2521 CATACTACCGGACGATCATCGGTGTTTACGCACGCTTCTCTTACGGTGCAGACT + 2580
 GTATGAGTCCCTGCATAGTAGCCACAATGCCGTGCAGAACGAGAAAATGCCACCGTCTGA
 C Y S R G R I I G V Y A R L A L Y G A D Y -
 2581 ACTTGATGCAAGAAAAGTAACGACTGGAATGCAATCAAAGAAATCGATGAAGAAACAA + 2640
 TGAACACTGCTCTTTTTCATTTGCTGACCTTACGTTAGTTCTTTAGCTACTTCTTTGTT
 C L M Q E K V N D W N A I K E I D E E T I -
 2641 TCCGTCCTCGTGAAGAAGTAACCTCAATACCAAGCATGCAACAAGTTGTTCCGCTGG + 2700
 AGGCAGAACCACTTCTTCATTTGGAAGTTATGGTTCGTAACGTTGTTCAACAAGCGGACC
 C R L R E E V N L Q Y Q A L Q Q V V R L G -
 2701 GTGACCTTACGGGTTGATGTTCCGAAACCAGGATGAACCTGAAGAGCAATCCAAAT + 2760

PCT/US05/27239 257/487

Figure 101F

CACTGGAAATGCCCAACTACAAGCGTTGGTCGCTACTTGCACCTTCTTCCTTAGGTTA
 D L Y G V D V R K P A M N V K E A I Q W -
 2761 GGGTTAACATTGCTTTTCATGGCTGTCTGCCGCTGTGATTAACGGTCTGTACATCTCTAG
 -----+-----+ 2820
 CCCAATTGTAACGAAAGTACCGACAGACGGCACACTAATGCCACGACGATGAGAGATC
 V N I A F M A V C R V I N G A A T S L G -
 2881 GTCCGTACC AATCGTATTGGACATCTTTGCAGAACCGTGACCTTCCTCGGTGACATTTA
 -----+-----+ 2880
 CAGCACATGGTTAGCATACCTGTAGAACCGTCTTGGCACTGGAACGAGCACCATGTAAT
 R V P I V L D I F A E R D L A R G T F T -
 2941 CTGAATCAGAAATCCAGAAATTCGTTGATGATTTGTTATGAACCTTCGTACAGTTAAAT
 -----+-----+ 2940
 GACTTAGTCTTTAGGTTCTTAAGCAACTACTAAGCAATACTTTGAAGCATCTCAATTTA
 E S E I Q E F V D D F V M K L R T V K F -
 3001 TTCCCTCGTACC AAGCTTATGACCAATTTGACTCAGGTGACCCACCTTTATCACAACCTT
 -----+-----+ 3000
 AAGGACATGGTTCCGAATACTGGTTAACATGAGTCCACCTGGGTTGGAAATAGTGTGAA
 A R T K A Y D Q L Y S G D P T F I T T S -
 3061 CTTATGGCTGGTATGGGTACGACGGTCTGTCACCGTGTACTAAGATGGACTACCGTTTCT
 -----+-----+ 3060
 GATACCGACCATACCCATTTGCTGCCAGCTGGCCACATGATTTCTACCTGATGGCAAGA
 M A G M G N D G R H R V T K M D Y R F L -
 3121 TGAACACTCTTGACAACATCGGTAACCTACCAGAACCAAACTTGACAGTTCTTTGGACTG
 -----+-----+ 3120
 ACTTGTGAGAACTGTTGTAGCCATTGAGTGGTCTTTGGTTTGAAGTGTCAAGAAAACCTGAC
 N T L D N I G N S P E P N L T V L W T D -

PCT/US05/27239 258/487

Figure 101G

```

3121 ~ACAAATGCCATACAACTCCGTCCTACTGTATGCACATGAGCCACAACACCTCTCTTA
+-----+-----+-----+-----+-----+-----+-----+-----+
3180 TGTTTAAGCGTATGTTGAAGGACCGGATGACATACGTTACTCGGTGTTGTGAGAAGAT
+-----+-----+-----+-----+-----+-----+-----+-----+
c K L P Y N F R R Y C M H M S H K H S S I -
+-----+-----+-----+-----+-----+-----+-----+-----+
3181 TCCAAATACGAAGGTGTAACAACAATGGCTAAAGACGGATATGGTGAATGAGCTGTATCT
+-----+-----+-----+-----+-----+-----+-----+-----+
3240 AGGTTATGCTTCCACATGTTGTACCGATTTCTGCCTATACCACCTTACTCGACATAGA
+-----+-----+-----+-----+-----+-----+-----+-----+
c Q Y E G V T T M A K D G Y G E M S C I S -
+-----+-----+-----+-----+-----+-----+-----+-----+
3241 CATGCTGTGTCCTCCACTTGATCCAGAAAATGAAGAACAACGCCACAACATCCAGTACT
+-----+-----+-----+-----+-----+-----+-----+-----+
3300 GTACGACACACAGAGGTGAACHTAGGCTTTTACTTCTTGTTCGGTGTGTAGGTCATGA
+-----+-----+-----+-----+-----+-----+-----+-----+
c C C V S P L D P E N E Q R H N I Q Y F -
+-----+-----+-----+-----+-----+-----+-----+-----+
3301 TCGGTCTCGTGTAAACGTTCTTAAAGCCCTTCTACTGGTGTGAATGGTGGTTACGACG
+-----+-----+-----+-----+-----+-----+-----+-----+
3360 AGCCACGACACATTTGCCAAGAAATTCGGGAAGAATGACCAAACTTACCACCAATGCTGC
+-----+-----+-----+-----+-----+-----+-----+-----+
c G A R V N V L K A L L T G L N G G Y D D -
+-----+-----+-----+-----+-----+-----+-----+-----+
3361 ATGTTCAAAAAGACTACAAAGTATTTGATATCGAACCAATCCGGTGAAGTCTTGAAT
+-----+-----+-----+-----+-----+-----+-----+-----+
3420 TACAAAGTGTTCATGATGTTTCATAAACTATAGCTTGGTTAGGCACCTCTTCAACAATTA
+-----+-----+-----+-----+-----+-----+-----+-----+
c V H K D Y K V F D I E P I R D E V L E F -
+-----+-----+-----+-----+-----+-----+-----+-----+
3421 TTGAATCAGTTAAAGCGAACTTTGAAAATCTCTTGACTGGTTGACTGACACTTACGGTAG
+-----+-----+-----+-----+-----+-----+-----+-----+
3480 AACTTAGTCAATTCGCTTGAACCTTTTATAGAACTGACCAACTGACTGTGAATGCATC
+-----+-----+-----+-----+-----+-----+-----+-----+
c E S V K A N F E K S L D W L T D T Y V D -
+-----+-----+-----+-----+-----+-----+-----+-----+
3481 ATGCCTTGAACATCATCCACTACATGATAGGTACAACTACGAAGCTGTTCAAATGG
+-----+-----+-----+-----+-----+-----+-----+-----+
3540 TACCGAACTTGTAGTAGGTGATGACTGACTATCCATGTTGATGCTTCGACAAAGTTTACC
+-----+-----+-----+-----+-----+-----+-----+-----+

```


PCT/US05/27239 259/487

Figure 101H

c A L N I I H Y M T D R Y N Y E A V Q M A -
 3541 CCTTCTGCCAACTAAACAACGGTCCCAACATGGGATTCGGTATCTGGGATTCGCTAACA + 3600
 GGAAGAAGGTTGATTTGTTGCACGGTTGTACCCCTAAGCCATAGACACCTAAACGATGT +
 c F L P T K Q R A N M G F I C G F A N T -
 3601 CTGTTGATACATTTGTCAGCTATCAAAATACGGTACAGTTAAACCAATCCGGTACGAAAGATG + 3660
 GACAACATGTAACAGTCGATAGTTTATGGGATGTCAAATTTGGTTAGGCACCTGCTTCTAC
 c V D T L S A I K Y A T V K P I R D E D G -
 3661 GCTACATCTACGATTACGAAACAATCGGTGACTACCCACCGCTGGGGTGAAGATGACCCAC + 3720
 CGATPAGATGCTAATGCTTTGTTAGCCACTGATGGGTGGACCCACCTTCTACTGGGTG
 c Y I Y D Y E T I G D Y P R W G E D D P R -
 3721 GTTCAAACGAATTGGCCAGAAATGGTTGATCGAAGCTTACACAACCTCGTCTACGTAGCCACA + 3780
 CAAGTTGCTTAAACCGTCTTACCACACTAGCTTCGAATGTTGAGCCAGATGCATCGGTGT
 c S N E L A E W L I E A Y T T R L R S H K -
 3781 AACTATACAAAGAGCCAGACTACAGTATCATCTTTGACAAATCATCTAACGTTGCTT + 3840
 TTGATATGTTTCTGGTCTTCGATGTCATAGTGAAAAGTGTAGTGTAGATTCCAACGAA
 c L Y K D A E A T V S L L T I T S N V A Y -
 3841 ACTCTAAACAACCTGGTAACTCACCCAGTTCACAAAAGGTGTATACCTCAACGAAAGATGGTT + 3900
 TGAGATTTGTTTGACCAATTGAGTGGTCAAGTGTTCACATATGGAGTTGCTTCTACCAA
 c S K Q T G N S P V H K G V Y L N E D G S -
 3901 CTGTGAACCTTGCTAAACTTGAATCTTCTCACCAGGTGCTAACCCATCTAACAAAGCTA + 3960

PCT/US05/27239 260/487

Figure 101I

GACACTTGAACACAGATTGAACTTAAGAAGAGTGGTCCACGATGGGTAGATTGTTGGAT
 V N L S K L E F S P G A N P S N K A K -
 AAGTGGTGGTTCACAAAACCTTGAACCTCAGTCTTCTAGCCCTTGACTTTACTTATGACAGCTG + 4020
 TTCCACCACCAACGTTTGAACCTTGAACCTTGAAGATCGGAATCGGAATCAATAACCGTCGAC
 C G G W L Q N L N S L S L D F S Y A A D -
 ACGGTATCTCATTGACTACACAAAGTATCACCTCGCGCTCTGGTAAAGACTCGTGATGAAC + 4080
 TGCCATAGAGTAACTGATGTTCATAGTGGAGCGGAGAACCATTCGTGAGCAGCTACTTG
 G I S L T T Q V S P R A L G K T R D E Q -
 AAGTTGATAACTTGGTAAACAATCCTTGATGGTACTTCCGAAAACGGTGGACAAACAGTTA + 4140
 TTCAACTANTGAACCATGTTAGGAACCTAGGAACCTACCAATGAAGCTTTTGCCACCTGTTGTGCAAT
 V D N L V T I L D G Y F E N G G Q H V N -
 ACTTGAACGTTATGGACTTGAACGATGTTACGAAAAAATCATGTCAGCGGAAAGAGTTA + 4200
 TGAACFTGCAATACCTGAACCTTGTCTACAAAATGCTTTTATAGTACAGTCCGCTTCTGCAAT
 L N V M D L N D V Y E K I M S G E D V I -
 TCGTACGTATCTCTGGATCTGTGTAAACACATAAATACCCTCACCAGAACAAAACCTG + 4260
 AGCATGCATAGAGACCTATGACACATTTGTGATTTATGGAGTGGAGTCTGTTTGTGAC
 V R I S G Y C V N T K Y L T P E Q K T E -
 AATTGACACAACGTTCTCCACGAAGTCTTTCAATGGATGACGCCCTTGGATGCAATTGA + 4320
 TTAACCTGTGTTGCACAGAAGGTGCTTCAAGAAAGTTACCTACTGCGGAACCTACGTAAC
 L T Q R V F H E V L S M D D A L D A L S -

Figure 101J

4321 GCTAATCAAGTCTTGAATAATAAAGGGCTCTTTGTCAACTGTAGTGGTGAAGAAA 4380
 -----+-----+-----+-----+-----+-----+-----+
 CGATTAGTTCAGAAGTCTTATTTTCCCGAGAAACAGTTGACATCACCACCAACTTCTTTT
 c *
 4381 AGCTAAGCTCGAGAAAGGACAAATTTGTCCTTTCTTTTGTGATTCAGAGCGATGAAA 4440
 -----+-----+-----+-----+-----+-----+-----+
 TCGATTTCGAGCTCTTTCCCTGTTTAAACAGGAAAGAAAAAATACTACAAGTCTCGCTACTTTT
 e * A R S L V F K T R E K K I N L A I F -orf1_670 homologue of sp0460, transposase
 4441 ATCCGTTTTTGAAGTTTTCAAAGTTCGGAAAAACCAAGGCAATTCGGCTTGATGCTTTTG 4500
 -----+-----+-----+-----+-----+-----+-----+
 TAGCAAAAAAATTCAAAAGTTTCAAGGCTTTTGGTTTCCGTTAACGGGAACACAGAAAAC
 e I R K K F N E F N R F G F A N R K I D K -
 4501 ATGAGTTTGTAGTGGCTCAAGTTTAGCGTTAGAAATAAGGCAATTCATGCGGTTAGTG 4560
 -----+-----+-----+-----+-----+-----+-----+
 TACTCAACAATCACGGGAGTTCAAATCCCAATCTATTTCCGTTAAGTTACCGCAATCAC
 e I L K N T A E L K A N S Y P L E I A N T -
 4561 ATGTAGTTTTTATAGCAATAAATGTCCTCAAAGTGGTTTTAAAGGTGCGGTTGAGATGA 4620
 -----+-----+-----+-----+-----+-----+-----+
 TACATCAAAAATATCGTTTATTTACACGAGTTTCACCAAAAATTTCCACGCCCAACTCTACT
 e I Y N K Y C I F T S L T T K F T R N L H -
 4621 GGTAACGTGCTTGAATTAAGCCCCAAAACCTGGTCAGTATCTCTCTTTGATGATGAAAT 4680
 -----+-----+-----+-----+-----+-----+-----+
 CCATTGCACAGAACTTAAATTCGGGTTTGACCAGTCATAGAAGAGAACATCTACTTTTA
 e P L T D Q I L G W F Q D T N K E Q L H F -
 4681 AGGAGTAGTGTACAGGTCAATGTAATCTTTAAGTTCAGGTTACTAGATAAGATTTTTC 4740
 -----+-----+-----+-----+-----+-----+-----+
 TCCTCATCAACTATGTCAGTATCATATAGAAAATTCAGTCCATGATCTCAATTTCTAAAAG

262/487

Figure 101K

e L L L Q Y L D Y Y D K L E P V L T F I K -
4741 TTGAGACACTCCCTAGGAGTTAAGGTCCTCTGAAAGTTCTAGCATAGAAAAGGCTTAAGA + 4800
AAGTCGTGAGGGATCCCTCAATCCAGAGACATTCAGAGATCCGATCTTCCGAAATTCI
e K L C E R P T L T E R F T R A Y F P K L -
4801 GAGAGTTCCGACTATCTTTTAGGATAAATTTCCAGTAATATTTAAGAGCCTCTGTATTCC + 4860
CTCTCAAAGGCTGATAGAAAATCCCTATTAAAGGTCANTATATAAATTCGAGACATAAGG
e S L K R S D K L I F K W Y Y K L A R Y E -
4861 AGAGATTTATGATCAAAATTCCTTCAAGATGTTGATTTCTAGTCTGATTAAGACCCCTGCTC + 4920
TCTCTAAATAGTAGTTAAAGGAGTACTACAACCTAAGATCAGACTAATTCCTGGGGACGAG
e L S K D D F Q K M I N I R T Q N L A R S -
4921 ATGTGTTGGACAATGTGGAAACGATCGAGAACAAATTTAGCATGGGAAATAATTTCTTA + 4980
TACACAACCTGTACACCTTTGCTAGCTCTGTTAAATCGTAACCTTTTATTAAGAAAT
e M H Q V I H F R D L V I K A N P F L K K -
4981 ATGAGAGGATATAACTTCCAGACAAATCAACAGTCAGACTTAACTTTTCTTCTAGCT + 5040
TACTCTCCCTATATTGAAGGCTCTGTATAGTTGTCATGCTGTAATGAAAAAAGATCGA
e I L P I Y S G S M D V T V V K V K K R A -
5041 TCTTTCCGAGTACTTGAAGAAATGATTCGGATGGTGTGTTGACGCTGTGTATCAAGAATG + 5100
AGAAAGCTCATGAACCTTCTTACTAAAGCCCTACCACAACTGCAGACAAATAGTTCTTAC
e E K S Y K F F H N R I T T Q R R N D L I -
5101 GTCATGATTTCTTAGTGTGAAATCCTGAGCAATGAAGCAATTCCTTCCCTTCTGCTAG + 5160

Figure 101M

5521 TTAGTGGGTGATAATCAAGTGTAGCCGAAAGACTTCGATATGGGTATCGTCTGAAATGGCT + 5580
 -----+-----+-----+-----+-----+-----+-----+-----+
 AATCCACCCACTATTAGTTCACATCGCTTCTGAAGCTATACCCATAGCACAGACTTACCGA

e K P P H Y D L T A F V E I H T D H Q I A -
 -----+-----+-----+-----+-----+-----+-----+
 5581 TTATTTAAGGTGATGTTTTTGGCTTTTATTCGGATGAGTAATGGTATGATGATGTT + 5640
 -----+-----+-----+-----+-----+-----+-----+
 AATAAATCCACTACAAAAACAGAAAAATAAGGCTACTCATTACACCATACTAATACACACA

e K N L T I N K D K I G I L L T T H N I H -
 -----+-----+-----+-----+-----+-----+-----+
 5641 TCCATAAGATACTTTCTAATGAGTTGTTTAGGGCTTTTTCATTTATAAGTCTTATGGGACT + 5700
 -----+-----+-----+-----+-----+-----+-----+
 AGGTATTCTATGAAAGATTACTCAACAAATCCCGGAAAAGTAATATTCAGATAACCCCTGA

e E M
 -----+-----+-----+-----+-----+-----+-----+
 5701 TTTTGTGATACTCAAAAAGCCCTAATAATCTCCAGAGTGGGATTTACCCACTACAGAAATTA + 5760
 -----+-----+-----+-----+-----+-----+-----+
 AAAAACTATGAGTTTTTTCGGGATATTAGAGGTGCACCCCTAAATGGGATGCTTTAAT

5761 TAGAGCCAGAAAAACACTTTTGTCTACTAGCAGAAACTAGAGAGCAGAAAGTGTTTTCT + 5820
 -----+-----+-----+-----+-----+-----+-----+
 ATCTCGGTCCTTTTGTGAAAAACAAGTGTGATCGTCTTTGATCTCTCGTCTCACAAAAAGA

5821 GTTCAGATTTACCCAAAACTGGGAAATATGGGGATAAGAATAGAGATGGCTTAGGAAGCC + 5880
 -----+-----+-----+-----+-----+-----+-----+
 CAAGTCTAAATGGGTTTGGACCCCTTATACCCCTATTTCTTATCTCTACCGAATCCCTTCGG

5881 CCTTTTGTGTGACAGTACCGATGAACCTTATAACAAAATAGTGAGCCCTTTTATGCAATC + 5940
 -----+-----+-----+-----+-----+-----+-----+
 GGAAAAACACACAGATCTGTCATGCTACTTGAATATTGTTTATCCTCCTCGGAAAAAATCGTTAG

e * L L Y H A K K A I -orf2_670 homologue of sp0461, transcriptional regulator
 -----+-----+-----+-----+-----+-----+-----+
 ATTCGACCCCGTTTGTCAAAAGCCCTTTTTCGGATATCTACAATTTGCTGATAGATGAGA

266/487

Figure 1010

```

6361  TCCTGAACAATTGCTTTCGAAATATGATACAGTGGCTTGTGCTTTCATCCCATTAATGTT+ 6420
-----+-----+-----+-----+-----+-----+-----+-----+
AGAACTTGTAAACGAAAGCTTTATATACTATGTCACCGACAGCGAAAGTTAGGGTATTACA
e   EQVIAKSIH Y L P K D S E I G Y H -

6421  TCGTAATAATTATAAGGAACTAGATTTTGTAAACCAACAAAACGTTCTTGTGTTAAG+ 6480
-----+-----+-----+-----+-----+-----+-----+-----+
AGCATTATTAATATATCCCTTGATCTAAAACATTTGGTTTGTGTTTGCAGAACAAATTC
e   EYNYYPV L N Q L G F L F T R T L -

6481  AAGTCAGTCTTAAAAAAGAAAGAAATTCGAAATGTCATTTCCCTAAGATATCTTTG+ 6540
-----+-----+-----+-----+-----+-----+-----+-----+
TTTCAGTCACGACAATTTTCTTCTTAAGCTTTACAGTAAAGGATTCCTATAAGAAC
e   FTLALFLSLSNSISIDNGLINK -

6541  AACTGGATAGTAGATGCTTTCCTCTTGTATGCTGAAGAATCAGTTGAATAGTATGAGTC+ 6600
-----+-----+-----+-----+-----+-----+-----+-----+
TTGAACCTATCATCTACGAAAGGAGAACATACGACTTCTTAGTCACTTATCATACTCAG
e   FKSLLLHKGRTHQLILQIHTH -

6601  TTTTTTCTTGATCCATTTGCTCCTTGGAAAACGAAATTAGCAGAACAAATAACCAA+ 6660
-----+-----+-----+-----+-----+-----+-----+-----+
AAAAAAGAACTAAGGTAACAGGAACCTTTGCTTCTTAATCGTCTTGTATTGTTGTTT
e   KKEQNWKKDKSFSNSASCVL -

6661  AAGATATAATCCAGTTCCTCTGAGTAAAGTCATGTTGGCATGTTGGCTCTAAGTAAGTT+ 6720
-----+-----+-----+-----+-----+-----+-----+-----+
TTCTATATTAGGTCAAGAAGGACTCATTTTCAGTACACCGTACACCGGAGATTCATTCAA
e   FIYDLEEQTF TMNAHP E L Y T -

6721  TGGAAATGTTCCATCAAAAICGGATACATAAAGAGGTTTTTAAATTTTCAAACTCTTTG+ 6780
-----+-----+-----+-----+-----+-----+-----+-----+
ACCTTACAAGGTAGTTTTAGCCTATGTAATTTCTCCAAAATAAAAAGTTTGAGAAAC

```


268/487

Figure 101Q

GGAAATCATCTTTCTCGCCATATAAAATTTTATCAACTAACCAATATTTTCGACTACC
e G Q L L F F R L I K L I T S Q N Y L Q H -
7201 AAGTAATAATTCGTTTGTGATGAGAATGGTTCGATTAATGAACCTGTTGCGTATCTAAA
TTCTATTTAAGCAAACTACTCTTACCACAGCTAATTAACCTTGAACAACGCATAGATTT
e F Y Y N T Q H S H E I L Q V Q Q T D L -
7261 TTAAATGTCAACTCTTCCTCGAATGTTTCTGTGTAATTCCTGCAAAATGCTTAGGAGACTT
AATTTACAGTTGAGAAGGAGCTTACAAGAACATTAAGGACGTTTACGAATCCTCTCGAA
e N F T L E E F T E Q L E Q L I S L L S -
7321 TTAGATTTGTAATGAAGTTAAAGTAGACAGTTCATCTAGTTCATATAGACCGAATATCCAAT
AATCTAACATTTACTTCAATTTCACTCTGCAAGTAGATCAAGTTAICTGCTTATAGGTTA
e K S Q L S T L T S L E D L E I S R I D L -
7381 AATATATTTAAAAGGGTAATTTTATCTGTAAATCTTTTTCATATGATTTGTTAGCATA
TTATATATAATTTTACCATTAAAATAGACATTAAGAAAAGTTACATAAACAAATCGTAT
e L I N L I T I K D T I R K E I Y K N L M -
7441 GTTACCGAATCTTAGTTGCATATAGATAATTTAATTAATATAAATACAAAAGAACTAAT
CAATGGCTTAGAATCAACGTTATCTATTAAAATTAATAATAATATGTTTTCTTTTGATTA
e T G T C T T G T C A A A A A G G T T G T G A A T T C C G A C T T T A T T G A T A A A A C A G C A T C T A A T A A A A
7501 A C A G A A C A G T T T T T C C A C A C C T T A A G G C T G A A A T A C T A T T T T G T C G T A C A T T A T T T T
e G G C A T T T T A A G A T A G T A T G A G T A T T G G T G G A G T T T A T G C C T A T T T T T T T A T T A T T A G A
7561

Figure 101R

CCGTAATAATTTCTATGCAATTAACCTCAATAACCCAAATAACCCGAATAAAAAAATAAATCT

AAATATTTTTTTTATCAAAATATGTCGTTCTATATAAAAAAATATGATGATAAAAAATATCTATT
-----+-----+-----+-----+-----+-----+-----+-----+
TTTATAAAAAAATAAGTTTATAACAGCAAGATATTTTTTTTATACACTATTTTTTATAGATAA

7680

GTGATGGAAGTGTGTTAAATTTATACTAGGATAGTAAATAGTAATACTATACATACTAT
-----+-----+-----+-----+-----+-----+-----+-----+
CACTACCTTCAACAAAATTAATATGATCCCTATGCAATTAATCAATATGATATGATATGATA

7740

ATTGTATACAAGTGTGTCATGCCAGTTCGAGAAGATAGCTATAACGCCACTTTTATACGC
-----+-----+-----+-----+-----+-----+-----+-----+
TAAACATATGTTCAACAGTAACGGTCCAACTCTTCTATCGAATATGCCGTGAAAATATGCCG

7800

TTTTGCTACGTTTGTTAGTGAACGGATTAACCTCAGTGAGATAAATTTTATCAGAACAATA
-----+-----+-----+-----+-----+-----+-----+-----+
AAAACGATGCAAAACAATCCTTGGCTAAATGAGTCACTCTATTTAAAATAAGTCTTTGTATT

7860

GTAATCCGTTTCTTCGTCATACAGATTGAAAGTACCTATGAAATCATAGAAGGATTAAC
-----+-----+-----+-----+-----+-----+-----+-----+
CATTAGGCAAGAAGCACATATGCTAACTTTTCATGGATACCTAGTATGCTCCCTAATTGA

7920

TGTTCTATGAATAATGCTTAAACAGGAGACACACATGAAAAAAGTAAAGATAATTCA
-----+-----+-----+-----+-----+-----+-----+-----+
ACAAGTACTTATTACGAATTGTCCTCTGHTGTACTTTTTCATTTCTTTCTATATAAGT

7980

GAAGGCAGTTGCAGGACTGTGCTGATATCTCAGTGCACAGCTTTTCTTCGATAGTTGC
-----+-----+-----+-----+-----+-----+-----+-----+
CTTCCGTCACGCTCCTGACAGCACATATAGAGTCAACTGTCCGAAAAGAACCTATCAACG

8040

b M L N R E T H M K K V R K I F Q -orf3_670 homologue of sp0462, LPXIG

b K A V A G L C C I S Q L T A F S S I V A -

270/487

Figure 101S

8041 TTATGAGAAACCGCCTGAAACCAGTCCAGCGATAGGAAAAGTAGTGATTAAGGAGACAGG + 8100
 -----+-----+-----+-----+-----+-----+-----+
 AAATCGTCTTTCCGGACTTTGGTCAGGTCGCCTATCCCTTTTCATCACTAAATTCCTCTGTCC
 b L A E T P E T S P A I G K V V I K E T G -
 CGAAGGAGGAGCGCTCTTAGGAGATGCCCTCTTTGAGTTGAAAAACAATACGGATGGCAC + 8160
 -----+-----+-----+-----+-----+-----+-----+
 GCCTCCCTCCCGGAAGATCCCTCTACGGCAGAAACCTCAACTTTTGTATTGCCTACCGTGG
 b E G G A L L G D A V F E L K N N T D G T -
 AACTGTTTCGCAAGGACAGAGCGCAACAGGAGAGCGGATATTTCAACATAAAACC + 8220
 -----+-----+-----+-----+-----+-----+-----+
 TTGACAAAAGGTTTCTGTCCTCCGGTTTGTCTCTTCCTTCCTGCTATAAAAAGTTTGTATTTTGG
 b T V S Q R T E A Q T G E A I F S N I K P -
 TGGACATACACCTTGACAGAACCCCACTCCAGTTGGTTATAAACCCCTACTATAACA + 8280
 -----+-----+-----+-----+-----+-----+-----+
 ACCCTGATGCGAACTGCTCTCCGGTTGAGGTCACCAATAATTTGGGAGATGATTTGT
 b G T Y T L T E A Q F P V G Y K P S T K Q -
 ATGACTGTTGAAGTTGAGAGAATGGTCGGACGACTGTCCAAGTGAACAGGTAGAAAA + 8340
 -----+-----+-----+-----+-----+-----+-----+
 TACCTGACAACCTCAACTCTTCTTACCAGCCTGCTGACAGGTTCCACCTTGTCCATCTTTT
 b W T V E V E K N G R T T V Q G E Q V E N -
 TCGAGAAGAGGCTCTATCTGACCAGTATCCACAAAACAGGGACTTATCCAGATGTCAAAC + 8400
 -----+-----+-----+-----+-----+-----+-----+
 AGCTCTTCTCCGAGATAGACTGGTTCATAGGTGTTTGTCCCTGAATAGGTCATACAAGTTTG
 b R E E A L S D Q Y P Q T G T Y P D V Q T -
 ACCTTATCAGATTATAAGGTAGATGGTTCCGAAAAAACAACGACAGCACAAAGCGTTGAA + 8460
 -----+-----+-----+-----+-----+-----+-----+
 TGGAATAGTCTAATAATTCATCTACCAAGCCTTTTGTTCCTGCTGCTGCTCCGCAACTT

PCT/US05/27239

271/487

Figure 101T

b P Y Q I I K V D G S E K N G Q H K A L N -
 TCCGAATCCATATGAACGGTGTGATTCAGAGGTCACACCTTCAAAAGAGAATTTATCAAGT + 8520
 8461 AGCTTAGGTATATCTCCACACTAAGGTCCTCCATGTGAAGTTCCTCTTAATAATAGTTCA
 P N P Y E R V I P E G T L S K R I Y Q V -
 GAATAATTGGATGATAAACCAATATGGAATCGAGTTGACGGTTAGTGTAATAAACGACGGT + 8580
 8521 CTTATTAAACCTACTATATGGTTATACCTTAGCTCAACTGCCAATCACCAGTTTGGCTGCCA
 N N L D D N Q Y G I E L T V S G K T T V -
 TGAACGAAAGAAGCCTCTACTCCGCTAGATGTTGTTATTCTATTAGATAACTCCAATAG + 8640
 8581 ACTTTGCTTCTTCGGAGATGAGCCGATCTACAACAATAAGATAATCTATTGAGGTTATC
 E T K E A S T P L D V I L L D N S N S -
 TATGAGTAATATTCGACATAATCATGCCCCATCGACGGGAAAAAGCGGAGAACGACACG + 8700
 8641 ATACTCATATATAAGCTGTATTAGTACGGGTAGCTGCCCTTTTTCGCCCTCTTCGGCTGTGC
 M S N I R H N H A H R A E K A G E A T R -
 AGCCCTTCTAGATAAGANTACCTCCAATCCAGATAATCGAGTAGCCTGTGACTTATGG + 8760
 8701 TCGGGAACATCTATTCTAATGGAGTTAGGTTAGTCTATTAGCTATCGTGAACACACTGAATACC
 A L V D K I T S N P D N R V A L V T Y G -
 CTCAACTATCTTTGACGGTTCAGAAGCTACTGTGGAAAAAGGGGTAGCAGATCGGAACGG + 8820
 8761 GAGTGTAGAAAACCTGCCAAGTCTCGATGACACCTTTTCCCCCATCGTCTACGGCTTGCC
 S T I F D G S E A T V E K G V A D A N G -
 AAAAATATGATGACTCAGCTTTATGGACGTTTCGATCGTTCAGACGTTTACAGCTAAAC + 8880
 8821

Figure 101U

8881 TTTTATAACTTACTGAGTCGAAATACCTCGCAAGCTAGCATGCTGCAAAATCGATTTTG
 b K I L N D S A L W T F D R T T F T A K T -
 TTATAATTAGAGCTTTTAAATCTCACATCAGATCCTCTACTGATATTCAAACTATTAAGGA 8940
 AATATTAATATCGAAAAATTTAGAGTGTAGTCTAGGATGACTATAAGTTTCATATAATTCCT
 b Y N Y S F L N L T S D P T D I Q T I K D -
 TAGGATTCCTCAGATGCAGAGGAATGAAACAAAGACAAAATGATGATCAANTCGGGCC 9000
 ATCCTAAGGTACTCTACGTCCTCTTAACTTGTTCCTGTTTAACTACATAGTTAAGCCGGC
 b R I P S D A E E L N K D K L M Y Q F G A -
 GACTTTTACCCAGAGGCTTTGATGACCGCTGATGATATCTTGACAAAAGCAGGCAAGACC 9060
 CTGAAAATGGGTCTTCCGAAACTACTGGCGACTACTATAGAACTGTTTCGTCCTGTTCTGG
 b T F T Q K A L M T A D D I L T K Q A R P -
 AAACAGTAAAAGGTTATTTCCACATACAGATGGTGTCCGACTATGTCATATCCAAAT 9120
 TTTGTCAATTTTCCAAATAAAGGTGTAATGCTTACCACAAAGGCTGATACAGTATAGGTTA
 b N S K K V I F H I T D G V P T M S Y P I -
 9121 TAAATTTAAATATACAGGAACGCGAATCGTACAGAACTCAGCTGAATATTTAAAGC
 ATTAATAATATATATGTCCTGCTGCTGCTAGCATGCTGAGTCTGACTTATAAAATTTCCG 9180
 b N F K Y T G T T Q S Y R T Q L N N F K A -
 AAAAATCCAAAATAGTAGCGGGATATTTACTGGAGGACTTTGTTACATGGTCAGCAGATGG 9240
 TTTTGGGTTTATCATCGCCCTATAATGACCCTCCCTGAAACAATGTACCAGTCGTCACCC
 b K T P N S S G I L L E D F V T W S A D G -

PCT/US05/27239

273/487

Figure 101V

9241 TGAACATAAGATTGTTCTGGTGGAGATGGTGAAGAAGTTATCAGANGTTTACGAAGAAACCTGT
-----+-----+-----+-----+-----+-----+-----+-----+ 9300
ACTTGTATCTTAACAAGCACCTCTACCCACTTTCAATAGTCTACAAATGCTTCTTTGGACA
E H K I V R G D G E S Y Q M F T K K P V -
9301 AACAGACCAATACCGGAGTTCATCAAAATACCTTTCAATCACCTCCATGGAGACAGAGACTAA
-----+-----+-----+-----+-----+-----+-----+ 9360
TTGCTGGTTATGCCCTCAAGTAGTTTATGAAAAGTTAGTGGAGGTACCTCGTCTCTCGATT
T D Q Y G V H Q I L S I T S M E Q R A K -
9361 ATTAGTTTCAGCGGGATATAGGTTCTATGGAACTGACTGCTTGTATTTTATTTGGCGTGATAG
-----+-----+-----+-----+-----+-----+-----+ 9420
TAATCAAAGTCGCCCTATATCCAAAGATACCTTGGACTGAACATAAATAAACCGCACATATC
L V S A G Y R F Y G T D L Y L Y W R D S -
9421 TATTCTAGCCCTATCCGATTTAACTCTAGTACCGATGGATTGATACCAACCATTGGTGACCCCTAC
-----+-----+-----+-----+-----+-----+-----+ 9480
ATAAGATCGGATAGGTAATAATTGAGATCATGGCTAACCTAARGTTGGTACCACCTGGGATG
I L A Y P F N S S T D W I T N H G D P T -
9481 GACTTGGTATTATACGGGAATATGGCTCAGGATGGCTATGATGTTCTTCACCTGTTGGGGT
-----+-----+-----+-----+-----+-----+-----+ 9540
CTGAACCATATAATTGCCCTTTATACGGAGTCCCTACCGATACAGAAAGTGACAACCCCA
T W Y Y N G N M A Q D G Y D V F T V G V -
9541 TGGTGTAAACGGGATCCTGGTACGGATGAAGCAACGGCTACTAGATTTATGCAGAGCAT
-----+-----+-----+-----+-----+-----+-----+ 9600
ACCACATTTGCCCTTAGGACCATGCTACTTTCGTTGCCGATGATCTAAATACGTTCTCGTA
G V N G D P G T D E A T A T R F M Q S I -
9601 CTCTAGTTCTCCGACAACTACATAACGGTAGGAGATCCATCTCAGATTTTACAAAGATT
-----+-----+-----+-----+-----+-----+-----+ 9660
GAGATCAAGAGGACCTGTGATGTGATTGCATCTCTAGGTAGAGTCTAAATGTTCTTAA

Figure 101W

b S S S P D N Y T N V A D P S Q I L Q E L -
9661 GAATCGCTACTTCTATATACTCGTCAATGAGAAGAAATCTATCGAAATAATGGTACGATTAC 9720
C TTAGCGATGAAGATATGATAGCAGTTACTCTCTTTTAGATAGCTTTTACCAATGCTAATG
b N R Y F Y T I V N E K K S I E N G T I T -
9721 AGACCCGATGGGGAACCTAATTGATTTCCAAATGGGAGAGATGGAGGTTTGTATCCAGC 9780
TCTGGCTACCCACTTGATTAAGTTAAAGTTAACCCCTCGTCTACCTCCAAACATAGGTCG
b D P M G E L I D F Q L G A D G R F D P A -
9781 GGATTACACTTTAACTGCAACGATGGTAGTTCGTTGGTCAATAATGTCCTCTCTGGGGG 9840
CCTAATGTGAATGACGTTTGTACCATCAAGCAACCCTTATTATACAGGGATGACCCCC
b D Y T L T A N D G S S L V N N V P T G G -
9841 ACCAATAATGATGGTGGCTTGTCTAAAAAATGCAAAAGTCTCTATGATACGACTGAGAA 9900
TGGTGTTTTACTACACCGAACGATTTTACTAGTTTTTCAAGATACTATGCTGACTCTTT
b P Q N D G G L L K N A K V F Y D T T E K -
9901 AAGGATCGGTGTAACAGGTTTGTACCTTGGAAACGGGTGAPAAAAGTTACATGACTTATAA 9960
TTCCTAAGCACATTTGTCCTCAACATGGAACCTTGGCCACCTTTTCAATGTAACATGATTT
b R I R V T G L Y L G T G E K V T L T Y N -
9961 TGTTCGCTGAATGACCAATTTGTAAGCAATAAATCTATGACACAGGAATGGTCGAACAAC 10020
ACAAGCGAACTTACTGGTTAAACATTCGTTAATTAAGATACTGTGCTTACCAGCTGTGTTG
b V R L N D Q F V S N K F Y D T N G R T T -
10021 CCTACACCCCTAAGGAAGTAGAAAAGAACACAGTGGCGGACTTCCCGATTCCCTAAGATTCCG 10080

Figure 101X

GGATGTGGGATCCCTTCATCTTTCTGTGTCACCGCTGAAGGGCTAAGGATCTAAGC
L H P K E V E K N T V R D F P I P K I R -
10081
TGATGTACGAAAGTATCCAGAAATCACAATCCAAAAGAGAAAACCTGGTGAATGA
ACTACATGCTTTCATAGGCTTTAGTCTTAAGGTTTCTCCTTTTGAACCACTTTAACT 10140
D V R K Y P E I T I P K E K L G E I E -
10141
GTTTATTAGATCAATAGAAATGATAAAAAACCCTGAGAGATGCCGCTTTAGTCTTCA
CAATAATCTAGTTATTCTTACTATTTTTGGTACTCTCTACGCCAGAAATCAGAAAT 10200
F I K I N K N D K K P L R D A V F S L Q -
10201
AAAACAACATCCGGATTATCCAGATATTATGGACTATTGATCAAAAATGGCACCTATCA
TTTTGTTAGGCCATAAGGCTATAAAATACCCTGATACTAGTTTTTACCGTGAATAGT 10260
K Q H P D Y P D I Y G A I D Q N G T Y Q -
10261
AAATGTGAAACAGGTGAAGATGGTAAGTTGACCTTTAAAATCTGTACAGATGGGAATA
TTTACACTCTGTCCACTTCTACCAATCAACTGGAAATTTTGTAGACAGTACCCCTTTAT 10320
N V R T G E D G K L T F K N L S D G K Y -
10321
TCGATTATTGAAAATCTGAACCCAGCTGGTTATAAACCCGTTCAAAAATAAGCCATTCGT
AGCTAATAAACCTTTTAAGACTTGGTCGACCAATATTGGGCAAGTTTTTATTCGGATACA 10380
R L F E N S E P A G Y K P V Q N K P I V -
10381
TGCCCTCCAAAATAGTAAATGGAGAGTCAGAGATGTGACTTCAATCGTTCACACAAGATAT
ACGGAAGTTTATCATTTACCTCTCAGTCTCTACACTGAAGTTAGCAAGGTGTTCTTATA 10440
A F Q I V N G E V R D V T S I V P Q D I -

Figure 101Y

10441 ACCAGCGGTTACGAGTTTACGAAATGATAAGCACTATATATCACAATAATGAGCCAAATTCCTCC
 -----+-----+-----+-----+-----+-----+-----+-----+
 10500 TGGTCCGCCAATGCTCAAATGCTTACTATATTCGTGATATAGTGTTTACTCCGGTTAAGGAGG

b P A G Y E F T N D K H Y I T N E P I P P -

10501 AAAAAAGAAATATCCTCGAACTGGTGGTATCGGAATGTTGCCATTCTATCTATGATAGGTTG
 -----+-----+-----+-----+-----+-----+-----+-----+
 10560 TTTTTCCTTATAGGAGCTTGACCACCAATAGCCTTACAAACGGTAAAGATAGACTATCCAAC

b K R E Y P R T G G I G M L P F Y L I G C -

10561 CATGATGATGGGAGGAGTTCTATATATATACACACGGAAACATCCGTAAGAAGTGTAGCAATGAG
 -----+-----+-----+-----+-----+-----+-----+-----+
 10620 GTTACTACTCCCTCACAAGATAATATGTGTGCCCTTGTAGGCCATTTCCACATCGTTACTTC

b M M M G G V L L Y T R K H P *

10621 AAATGATAAATATCGATACTCTGAGCGGATACTTTTAAAGAAAGTAGCACTCAAGAAGAGATTT
 -----+-----+-----+-----+-----+-----+-----+-----+
 10680 TTTACTATATAGCTATGAGACTCGCTATGAAATTTCTTCATCGTGAGTTCTCTCTTAAA

10681 AAGTTTACTTGGTGAACAGTTTCTTCGCCAAGTAAACCACCATTGAAGGGGAGATG
 -----+-----+-----+-----+-----+-----+-----+-----+
 10740 TTCAAATGAACCACCTTTGTCAAAGAAAGCGGTTCAATTTGGTGGTAACTTCCCTCTAC

10741 TTTTTCGAAACTGCACAGAAAAAGGATATATTTGTCATGTGTAATTCATTACATTGC
 -----+-----+-----+-----+-----+-----+-----+-----+
 10800 AAAAGCTTTTGAACGGTGTCTTTTTTCCATAATAAACAAGTACACATTAAGTAATGTAACG

10801 TCACAGTTGATTTAAGAGATATGAAATAGGAGAAATCATGAATCAATCAACAAATTTT
 -----+-----+-----+-----+-----+-----+-----+-----+
 10860 AGTGTCAACTAAAATTCCTATACTTATTCCTCTTTTACTACTTTTACTTTAGTTGTTTAAAA

c TAACAAATGCTTGTGCTCTTATTACTGACACGGAGTAGCCTGTTTTCAGCTGCAACAGTTT M K S I N K F L - orf4_670, homologue of sp0463, LPXTG

PCT/US05/27239

278/487

Figure 101AA

11281 TTGGTCCTTAATGGGCAAGTAAATAACAGTTCAAAGCGGTACCTGCTCTTGTAACTCTTC + 11340
 AACCAGGATTACCCGGTTCATAAATGTCACAGTTTCCAAAGTTTTCGGCATGGACGAGAACATFGAGAAG
 C G P N G Q V L T G S K A V P A L V T L P -

11341 CACTTGTAAACAATAATGGTACAGTAATGATGCACATGTTTCCCTAAAAAATTCATATA + 11400
 GTGAACAATTGTTATACCAATGTCATTAACACTACGGTACAAAAGGGATTTTAAAGTATAT
 L V N N N G T V I D A H V F P K N S Y N -

11401 ATAAACCAGTTGTAGATAAAAAGAATGCTGATACITTTGAATTAATAACGATCAAAAATGGTC + 11460
 TATTTGGTCAACATCTATTTTCTTAACGACTATGAACCTTAATATGCTAGTTTACCAG
 K P V V D K R I A D T L N Y N D Q N G L -

11461 TGCTATCGGTACTAAAATCCCATATGTTGTTAATACAAACAATTCCAAGTAATGCAACAT + 11520
 AGAGATAGCCATGATTTTAGGGTATACAAACAATTAATGTTGTTAAGGTTCAATACGTTGTA
 S I G T K I P Y V V N T T I P S N A T F -

11521 TTGCAACTTCATTTGGTCAGATGAAATGACAGAGGCTTAACCTTATAATGAAGATGATA + 11580
 AACGTTGAAGTAAACCAGTCTACTTTACTGTCTCCAGATTGAAATATTACTTCTACAT
 A T S F W S D E M T E G L T Y N E D V T -

11581 CAATTACTTTGAATAATGTAGCTATGGATCAAGCTGATTAATGAAGTCACTAAAAGGAATA + 11640
 GTTATGAACACTTATTACATCGATACCTAGTTCGACTAATACTTCAGTGAATTCCTTTAT
 I T L N N V A M D Q A D Y E V T K G N N -

11641 ATGGCTTTAACTTAAAATTAACAGAACGAGTTTAGCTAAAATAATGTAAGGATGCGAG + 11700
 TACCGAAATGAATTTTAAATGTTCTCGTCCAAATCGATTTTAAATTAATACCATTCTACGTC

Figure 101AB

C G F N L K L T E A G L A K I N G K D A D -
 ACCAAAAATCCAAATTACTACTCAGCTACTTTGAACACTCCTTGCTGTTGCAGACATTC 11760
 TGGTTTTTAGGTTTAAATGAATGAGTCGATGAAACTTGAAGTGAACCGACAACGGTCTGTAAAG
 C Q K I Q I T Y S A T L N S L A V A D I P -
 CTGAAAGTAACGATATTACATATCATTTACGGAAATCATCAAGATCATGGGAATACATCCAA 11820
 GACTTTCATTCCTATAATAGTATAGTAATGCGCTTTAGTAGTTCTAGTACCCCTTATGAGGTTT
 C E S N D I T Y H Y G N H Q D H G N T P K -
 AACCAACTAAACCTAATAATGGTCAATTTACAGTAACCTAAGACATGGGACAGTCAACCTG 11880
 TTGGTTGATTTGGATTAATACCAGTTTAAATGTCATTAATCTGTACCCCTCAGTTGGAC
 C P T K P N N G Q I T V T K T W D S Q P A -
 CTCCTGAGGGGGTGAACCGACTGTTCAACTTGAATAATGCCAAGACTGGTGAGAAAAGTCG 11940
 GAGGACTCCCCACTTTCGGCTGACAAGTTGAACATTTACGGTCTGACCACCTCTTTTCAGC
 C P E G V K A T V Q L V N A K T G E K V G -
 CTGCTCCCTGTAGAACTTTCAGAAAATAATTTGGACATATACTTTGGAGTGGTCTAGATAATT 12000
 CACGAGGACATCTTGAAGTCTTTTATTAACCTGTATATGAACCTCACCCAGATCTATTAA
 C A P V E L S E N N W T Y T W S G L D N S -
 CTATTGAATACAAAAGTTGAAGAAGAAATATAATGGATACTCAGCTGAATACACAGTAGAGA 12060
 GATAACTTATGTCTCAACTCTTCTTATATACCTATGAGTCGACTTATGTGTCATCTCT
 C I E Y K V E E E Y N G Y S A E Y T V E S -
 GCAAAGGGAAGTTGGGGTAAAAAACTGGAAAGATAATAACCCAGCTCCCAATCAATCCTG 12120

280/487

Figure 101AC

```

CGTTCCCTTCAACCCCAATTTTGGACCTTCTATATATGGGTCGAGGTTAGTTAGGAC
  K G K L G V K N W K D N N P A P I N P E -
12181 AAGAACACGGTGTAAACATACGGTAAAGTTTGTCAAGTAGACCAAAAAGATACTC
TTCCTGGTGCACATTTTGTAGCCATTTTCAACACAGTTTCATCCTGGTTTTTCTATGAG
  E P R V K T Y G K K F V K V D Q K D T R -
12181 GTCTAGAAAATGGCGAGTTCGTTGTTTAAAAAGCAGATACCAATAATATATCCCTTTA
CAGATCTTTTACGGCTCAGCAACAATTTTTTGGTCTATCGTTATTTATATAAGGAAAT
  L E N A Q F V V K K A D S N K Y I A F K -
12241 AGTCAACTGCACAACAAGCTGCAGATGAAAAAGCAGCAGCAACTGCAAAAACAATAATTGG
TCAAGTTGACGTTGTTGTCGACGCTACTTTTTTGGTCTGCTGCTGACGTTTTGTTTTAACC
  S T A Q Q A A D E K A A A T A K Q K L D -
12301 ATGACGGGTAGCAGCTTACAAAATGCTGCAGATAGCAAGCCGCTCAAGCTCTAGTAG
TACGTTGCCCATCGTCCAATGTTTACGACGCTATTTCTGTTCCGGGAGTTCGAGATCATC
  A A V A A Y T N A A D K Q A A Q A L V D -
12361 ATCAACACAGCAAGAATACAAATGTTAGCTTACAAAGAACCCAAATTTGGTTATGTTGAAG
TAGTTTCGTTGCTTCTTATGTTTACATCGAATGTTTCTTCGGTTTTAAACCAATACAACTTC
  Q A Q Q E Y N V A Y K E A K F G Y V E V -
12421 TAGCTGAAAAGATGAAGCAATGGTCTCTTACTTCTAATAACGGATGGTCAATTCGAAATTT
ATCGACCTTTTCTACTTCTGTTACCAAGAATGAGATTATGCTTACCTACAGGTTAAGGTTTAAA
  A G K D E A M V L T S N T D G Q F Q I S -

```

PCT/US05/27239 281/487

Figure 101AD

12481 CAGGCTTGGCTGGTACTTATAAATAAGAAATAAAGCTCCAGAACGTTTGGCGA 12540
 +-----+-----+-----+-----+-----+-----+
 GTCCAGAACGACGACCCATGAATATTTAAATCTCTTTAATTTCCGAGGTCCTCCAAAACGCT

12541 GLAAGTYKLEEKAP EGF AK -
 +-----+-----+-----+-----+-----+-----+
 AAATGATGATGATGATAATTTGTTGGAGCAGGTTCTGGAAATCAAGGTCAGTTTAAT
 12600
 +-----+-----+-----+-----+-----+-----+
 TTTAACTACTACATCTTAAACAACAACCTCGTCCAAAGAACCTTAGTCCACTCAAATTA

12601 IDDEV FV V G A G S W N Q G E F N Y -
 +-----+-----+-----+-----+-----+-----+
 ACTTAAAGATGTTCAAAGAATGACGCTACAAAAGTAGTCAACAAAAAATCACTATCC
 12660
 +-----+-----+-----+-----+-----+-----+
 TGAATTTCTACAAGTTTCTTACTCGCATGTTTTCATCAGTTGTTTTTTTAGTGATAG

12661 L K D V Q K N D A T K V V N K K I T I P -
 +-----+-----+-----+-----+-----+-----+
 CACAAAACGGTGGTATTGGTACAATTAATCTTTGCTGTAGCGGGGGCTGGATATGGGTA
 12720
 +-----+-----+-----+-----+-----+-----+
 GTGTTGCCACCACATAACCATGTTAATAGAAAACGACATGCCCGGACGGTAATACCCAT

12721 Q T G G I G T I I F A V A G A A I M G I -
 +-----+-----+-----+-----+-----+-----+
 TTGCAGTGTACGCATATGTTAAACAAACAAGATGAGGATCAACTTGCITTAAGTAAGAG
 12780
 +-----+-----+-----+-----+-----+-----+
 AACGTCACATCCGTTATACAATTTTGTGTTTCTACTCTCTAGTTGAACGAATTCATTCTC

12781 A V Y A Y V K N N K D E D Q L A *
 +-----+-----+-----+-----+-----+-----+
 AGAAAGGAGCCATTGATGACAATGCAGAAAATGCAGAAAATGATTAGTCGTATCTTTT
 12840
 +-----+-----+-----+-----+-----+-----+
 TCTTTCCTCGGTAAC TACTGTACGTCGTTTACGTCGTTTACGTCGTTTACGTCGTCATAGAGAAA

12841 M T M Q K M Q K M I S R I F F -orf5_670, homologue of sp0464, LPXTG
 +-----+-----+-----+-----+-----+-----+
 GTTATGGCTCGTGTTTTTCTCTTGTATGGGGTGCACATGCAGTCCCAAGCCCAAGAAGAT
 12900
 +-----+-----+-----+-----+-----+-----+
 CAATACCGAGACACAAAAGAGAACATACCCACCGTACGTCAGGTTCCGGTCTCTCTA

Figure 101AE

```

a V M A L C F S L V W G A H A V Q A Q E D -
12901 CACACGTTGGTCTTCAATTGGAGAACTATCAGGAGGTGGTTAGTCAATGGCCATCTCGT
-----+-----+-----+-----+-----+-----+-----+-----+
GTGCGCAACCAGAACGTTAACTCTTGTAGTCCCTCCACCAATCAGTTAACGGTAGAGCA
-----+-----+-----+-----+-----+-----+-----+-----+
12960
a H T L V L Q L E N Y Q E V V S Q L P S R -
12901 GATGGTCATCCGGTTGCAAGTATGGAAGTGGATGATTCGTATTCCTATGATCATCGGGTG
-----+-----+-----+-----+-----+-----+-----+-----+
CTACAGTAGCCAACGTTTCATACCTTCAACCTACTAAGCATAAGGATACTACTAGCCCCAC
-----+-----+-----+-----+-----+-----+-----+-----+
13020
a D G H R L Q V W K L D D S Y S Y D D R V -
13021 CAAATTGTAAGACACTTGCATTCGGATGAGAAATAAACTTCTCTTCAAAAAAGACT
-----+-----+-----+-----+-----+-----+-----+-----+
GTTTACATTCCTGACCGTAAGCACCCCTACTCTTATTGAAAGAAAGTTTTTCTGGA
-----+-----+-----+-----+-----+-----+-----+-----+
13080
a Q I V R D L H S W D E N K L S S F K K T -
13081 TCGTTGAGATGACCTTCCTTGAGAAATCAGATGGAATGAAATCTCATATCCAAATGGTCTT
-----+-----+-----+-----+-----+-----+-----+-----+
AGCAAACCTCTACTGGAGGAACTCTTAGTCTAATCTCATAGAGATTAAGGTTTACCAGAA
-----+-----+-----+-----+-----+-----+-----+-----+
13140
a S F E M T F L E N Q I E V S H I P N G L -
13141 TACTATGTTCCGCTTATATCCAGACGGATCCGGTTCTTATCCAGCTGAATTCCTTTT
-----+-----+-----+-----+-----+-----+-----+-----+
ATGATACAAGCGAGATAATAGGTCCTCCACGCCCAAGAATAGGTCGACTTAAAGAAAAA
-----+-----+-----+-----+-----+-----+-----+-----+
13200
a Y Y V R S I I Q T D A V S Y P A E F L F -
13201 GAAATGACAGATCAAACGGTAGAGCCTTTGGTCAATGATGCGAAAAAACAAGATACAATG
-----+-----+-----+-----+-----+-----+-----+-----+
CTTTACTGCTACTTGGCATCTCGGAACACAGTACATCGCTTTTGTGTATGTTAC
-----+-----+-----+-----+-----+-----+-----+-----+
13260
a E M T D Q T V E P L V I V A K K T D T M -
13261 ACACAAAGGTGAAGCTGATAAAGGTGGATCAGACCACAATCGGTTGGAGGCTCGGC
-----+-----+-----+-----+-----+-----+-----+-----+

```


Figure 101AF

```

a      T T K V K L I K V D Q D H N R L E G V G -
      TGTTGTTCCACTTCGACTATTTCCACCTAGTTCTGGTGTAGCGAACCTCCCACAGCCG
      TTTAAATTTGGTATCAGTAGCAAGATGGTTCTGAAAAAGAGGTTCCCTTGATTGGAGAA 13380
      AAATTAACCATAGTCATCGTTCTCTACCAAGACTTTTTCACAAGGGAACATAACCTCTT +
a      F K L V S V A R D G S E K E V P L I G E -
      TACCGTTACAGTTCCTTCTGGTCAAGTAGGGAGAACTCTCTATACTGATATAAATGGAGAG + 13440
      ATGGCAATGTC AAGAGACCAAGTTCATCCCTCTTGAGAGATATGACTATTTTTACCTCTC
a      Y R Y S S S G Q V G R T L Y T D K N G E -
      ATTTTTGTGACAAATCTTCCCTCTGGGAACATACGTTTCAAGGAGGTTGGAGCCACTGGCA + 13500
      TAAAACACTGTTTAGAAGGAGAACCTTGATAGCAAGTTCCTCCACCTCGGTGACCGT
a      I F V T N L P L G N Y R F K E V E P L A -
      GGTATGCTGTTACGACGCTGGATACGGATGTCACCGCTGGTAGATCATCAGCTGGTGACG + 13560
      CCGATACGACAATGCTGGACCTATGCTACAGGTCGACCATCTAGTAGTCGACCACTGC
a      G Y A V T T L D T D V Q L V D H Q L V T -
      ATTACGGTTGTC AATCAGAAATTTACCACCGTGGCAATGTTGACTTTATGAAGGTGGATGGT + 13620
      TAATGCCAACAGTTAGTCTTTAATGGTGCACCGTTACACTGAATAACTTCCACCTACCA
a      I T V V N Q K L P R G N V D F M K V D G -
      CGGACCAATACCTCTCTTCAAGGGCAATGTTCAAAAGTCATGAAGAAGAAAGCCGGACAC + 13680
      GCCTGGTATGGAGAGAAGTTCCCGTTACAAAGTTTCAGTACTTTCTTCTTCGCGCTGTG
a      R T N T S L Q G A M F K V M K E E S G H -

```


Figure 101AH

14101 TCACAATAACGGGATATGAGGCTGGCAGATGTGCCAGCCCTCATGTGGGTATTGTTTG + 14160
 -----+-----+-----+-----+-----+-----+-----+
 14161 AGTGTATTGCCCTATACCTCGGACCCGCTCTAACACGGTGGGAGTAACACCCCAATAACAAC

14221 TAAAAAGATAGGACTGGTCTGGTAATCATTTTAGGAATGGACAGACTGGGATCTCGATT + 14280
 -----+-----+-----+-----+-----+-----+-----+
 14281 AATTTGCTATCCTGACCCAGACCATTAAGTAAATCCTTACCCTGCTGACCCCTAAGACTAA

14341 TAAAAATGGATGGTGAATCAGAAAGAAATGAGATTTTCTCGTTTCTCTTAGCAGATAGGAT + 14400
 -----+-----+-----+-----+-----+-----+-----+
 14401 AATTTACCTACCACCTTAGCTCTTCTTTACTCTAAAGAGCAAGAGAAATCTCTATCCTA

14461 TCTCTGTTAGGAAAAGCGATATAAATGATGAGTTTGAAGATAAAGGGATGCTGATAAAAAT + 14520
 -----+-----+-----+-----+-----+-----+-----+
 14521 ACAGACAAATCCTTTTTCGCTATTTTACTACTCAAACTTCTATTTCCCTACGACTATTTTTA

M L I K M -orf6_670, homologue of sp0466, sortase

14581 CGTAAAAACAAAAAGCAAAAACGAAATAATCTCCTATATTAGGACTGGTATTTTCATTGG + 14640
 -----+-----+-----+-----+-----+-----+-----+
 14641 CCATTTTGTGTTTTTTCGTTTTTTCGTTTTTCTTTATTAGAGGATAATCCTCACCATAAAAAGTAACC

b V K T K K Q K R N N L L L G V V F F I G -

14701 AATGGCGGTAATGGCGTATCCCGTGGTCTCGCTTCTATATCGAGTGGAAATCAAATCA + 14760
 -----+-----+-----+-----+-----+-----+-----+
 14761 TTACCGCCATACCGCATAGCGGACCCACAGAGGGAACATAATAGCTCACCCTTAGTTTAGT

b M A V M A Y P L V S R L Y R V E S N Q -

14821 ACAAAATGCTGACTTTGATAAGGAAAAACCAACGTTGGATGAGGCTGACATGATGAACG + 14880
 -----+-----+-----+-----+-----+-----+-----+
 14881 TGTTTAACGACTGAAACTATTCCTTTTTCGTTGCAACCTACTCCGACTGTAACTACTTTC

b Q I A D F D K E K A T L D E A D I D E R -

AATGAAATGGCACAAGCCTTCAATGACTCTTTGAAATAATGATGAGTGGCGGATCCTTG

PCT/US05/27239

287/487

Figure 101AJ

14941 GATTGTACCAGGTCATGATTATGTCACCTTGGCTGACTTGTAGCCCATACATGATCAATAC + 15000
 CTAACATGGTCCAGTACTAATACTACTGGAACCGACTGAACATGCGGTATGACTAGTTATG
 b I V P G H D Y V T L L T C T P Y M I N T -

15001 CCATCGTCTATTGGTTCGGGGCATCGGATACCGTACGTTACGAGAGGTTGAGGAAGAATT + 15060
 GGTAGGAGATAACCAAGCCCCCGTAGCCTATGGCATGCACTCGTCTCCAACTCCTTCTTAA
 b H R L L V R G H R I P Y V A E V E E F -

15061 TATTGCAGCAAACTCAGTCACTCTATCGCTACCTGTTTATGTCGAGTTGGTTT + 15120
 ATAACGTCGTTTGTGAGTCAGTAGAGATAGCGATGGACAAAATACACCGTCAACCAAA
 b I A A N K L S H L Y R Y L F Y V A V G L -

15121 GATTGTGATCTTTTATGGATTATTCGACCGCTTCGCGAAGAAAGAAAACAACCGGAAA + 15180
 CTAACACTAAGAAAATACCTAATAAGCTGCGAAGCTGCTCTCTTTTGTGGCTTTT
 b I V I L L W I I R R L R K K K Q P E K -

15181 GGCTTTGAAGGGCTGAAAGCAGCAAGGAAGGAAGTGAAGTGGAGGATGGACAACAGTA + 15240
 CCGAAACTCCGGACTTTCGTCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT
 b A L K A L K A A R K E V K V E D G Q Q * -

15241 GACGTTACGAAAAAAGGCACAAAAAAGAAAGAAACATCCCGTGTATCCCTTCTTCGATTT + 15300
 CTGCAAGTCTTTTTCGGTGTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 b A L K A L K A A R K E V K V E D G Q Q * -

15301 TCCTAGTAGGATTCGGCGTTGGGATATATCCATTTGGTCTCGTTATATATATTCGTTATTTG + 15360
 AGAATCATCTAAGCGCAAGCCTATATAGGTAACCCACAGACAAATAATAATAGCATAAC
 c V S R Y Y R I E -orf7_670, homologue of sp0467, sortase

288/487

Figure 101AK

15361 AGTCAAACGAGGTTATTAAGAGTTTGAATGAGACGGTTTCCAGATGGATAAGGCAGAAC + 15420
 -----+-----+-----+-----+-----+-----+-----+-----+
 TCAGTTTCTCCCAATAATTTCACAACTACTCTGCCAAAGGGTCTACCTATTCCTCGTCTTG
 C S N E V I K E F D E T V S Q M D K A E L -
 15421 TTGAGGAGCGTTGGCGCTTGGCTCAAGCCTCAATGCGACCTTGAACCACTGTGAATTC + 15480
 -----+-----+-----+-----+-----+-----+-----+-----+
 AACCTCTGCCAACCCGGAACCGAGTTCGGAAAGTTACGGTGGAACTTGGTAGACTTTTAAG
 E E R W R L A Q A F N A T L K P S E I L -
 15481 TTGATCCCTTTACAGAGCAAGAGAAAAGAAAAGCCCTCTCAGAAATATGCCAATATGCTAA + 15540
 -----+-----+-----+-----+-----+-----+-----+-----+
 AACTAGGAAAATGTCCTGTTCTCTTTCTTTCCGAGAGCTTTATACGGTTATACGATT
 D P F T E Q E K K K G V S E Y A N M L K -
 15541 AGGTCCATGAGCGGATTTGGTATATGTGGAATTCCTCGGATGATCAGGAAATTCGGATGT + 15600
 -----+-----+-----+-----+-----+-----+-----+-----+
 TCCAGGTACTCGCCTAACCGATACACCTTTAAGGACGCTAAGTCTTTAAGGCTACA
 V H E R I G Y V E I P A I D Q E I P M Y -
 15601 ATGTCCGAACGAGTGAGGAAATTCCTTCAGAAAGGCGCCAGGATTCCTAGAGGGAGCTTCGT + 15660
 -----+-----+-----+-----+-----+-----+-----+-----+
 TACAGCCTTGCTCAGTCTCTTTAAGAAGTCTCCCGGCTTACCGAFCCTCCCTCGAAGCA
 V G T S E E I L Q K G A G L L E G A S L -
 15661 TACCGGTTGGTGGTGAATAATACCCACACAGTTGTCTCAGTCTCATAGAGGATTCACGACGG + 15720
 -----+-----+-----+-----+-----+-----+-----+-----+
 ATGGCCAACCACTTTTATGGGTGTGTCAACAGTACAGGATCTCCTAATGGCTGCC
 P V G G E N T H T V V T A H R G L P T A -
 15721 CAGAAGTGTTTAGTCAATTTGGATAAGATGAAAAAGGGGATGCTTTTATCTTCACGTTT + 15780
 -----+-----+-----+-----+-----+-----+-----+-----+
 GTCCTTGACAAAATCAGTTAACCTATTCCTACTTTTTTCCCTACAGAAAATAGAACTGCAAA

PCT/US05/27239 289/487

Figure 101A

c E L F S Q L D K K K G D V F Y L H V L -
 15781 TAGACCAGTGTGGCCCTACCAAGTGGATCAGATTTTGACGGTTGAGCCAAATGACTTTG + 15840
 ATCTGGTCACAACCGGATGGTTCACCTAGTCTATAAAGCTGCCAACCTCGTTTACTCTGAAC
 c D Q V L A Y Q V D Q I L T V E P N D F E -
 15841 AGCCTGTCTGATTCACATGGGGAAGATATGCGACCTTGTGTGACCTGTACACCCGTATA + 15900
 TCGGACAGAACTAAGTTGTACCCCTCTAATACCTGGAACAACCTGGACATGGGCATAT
 c P V L I Q H G E D Y A T L L T C T P Y M -
 15901 TGATTAACAGTCATCGTCTGTGGTACGTTGGGAAGCGGATTCGGTATACGGCCACCATTG + 15960
 ACTAATTCAGTAGCAGACAAACCATGCACCCTTCGCCCTAAGGCATATGCCGTGGTTAAC
 c I N S H R L L V R G K R I P Y T A P I A -
 15961 CAGAGGAAATCGAGCGGTGAGAGCGTGGGAAATTCGTGGTGTGGTTATGCTACCGG + 16020
 GTCTCGCTTAGCTCGGCACCTCTCGCACCCCTTAGACCAACACCAATACCGATGGCC
 c E R N R A V R E R G Q F W L L L L A A -
 16021 CGTTGGTTATGATTCCTGGTATTGAGTTACGGGGTGTATCGTCACTCGTCCGATTTGTAAG + 16080
 GCAACCAATACTAAGACCATACTCAATGCCACATAGCCACATAGCAGTAGCAGCGTACACATTC
 c L V M I L V L S Y G V Y R H R R I V K G -
 16081 GGCTAGAAAACAAATTGGAGGAGCATCATGTCAAAGGCTAAGCTACAGAAATTAAGGGG + 16140
 CCGATCTTTTGTAACTCCCTCGTAGTACAGTTCCGATTCGATGTCCTTAATGATCCC
 a M S K A K L Q K L L G -orf8_670, homologue of sp0468, sortase
 c L E K Q L E E H H V K G *

Figure 101AM

```

16141 TATTTGCTGATGCTGGTAGCATTGGTAGTTCCTGTTTATTTGTTTGGCAGATGGTGTTA 16200
-----+-----+-----+-----+-----+-----+-----+-----+
16142 ATAAACGACTACGACCATCGTAACCACTAAGGACAATAACAAAACCCGTTACTCCACAAT
a Y L L M L V A L V I P V Y C F G Q M V L -
-----+-----+-----+-----+-----+-----+-----+-----+
16201 CAGTCTTTAGGACAAGTAAAAGGTCATGAGATATTTTCAGAAATCTGTGACGGCCCGACAGT 16260
-----+-----+-----+-----+-----+-----+-----+-----+
16202 GTCAGAAAATCCTGTTTCATTTCCAGTACTCTATAAAAAGTCTTAGACACTGCCCGGCTGTCA
a Q S L G Q V K G H E I F S E S V T A D S -
-----+-----+-----+-----+-----+-----+-----+-----+
16261 TACCAAGACCAATTGCAACGGTTCGTTGATTACAAATCAACGGTTGGATTGCGCAAAATCGT 16320
-----+-----+-----+-----+-----+-----+-----+-----+
16262 ATGGTCTCGTTAACGTTGCCAGCGAACTAATGTTAGTTGCGAACCTAAGCGTTTTFAGCA
a Y Q E Q L Q R S L D Y N Q R L D S Q N R -
-----+-----+-----+-----+-----+-----+-----+-----+
16321 ATTGTAGATCCTTTTGGCGGAAGGTATGAGTAAATACCAAGTCTGTGACGATCCT 16380
-----+-----+-----+-----+-----+-----+-----+-----+
16322 TAACATCTAGGAAAAAACCCGCTTCCCATCTCCATTTAATGGTTCACAGACTGCTAGGA
a I V D P F L A E G Y E V N Y Q V S D D P -
-----+-----+-----+-----+-----+-----+-----+-----+
16381 GATGCAGTCTACGGCTATTTGTCGATTCGAGTTTGGAAATCATGGAGCCAGTTTATCTA 16440
-----+-----+-----+-----+-----+-----+-----+-----+
16382 CTACGTCAGATGCCGATAACAGCTAAGGCTCAAACCTTTAGTACCTTCGGTCAAATAGAT
a D A V Y G Y L S I P S L E I M E P V Y L -
-----+-----+-----+-----+-----+-----+-----+-----+
16441 GGAGCGGATTACCATCATTTAGCAATGGGTTGGCCCATGTTGGATGGACGCCCTTCCT 16500
-----+-----+-----+-----+-----+-----+-----+-----+
16442 CCTCGCTAATGGTAGTAATCGTTACCCCAACCGGGTACACCTACCCCTGCGGAGAAGGA
a G A D Y H H L A M G L A H V D G T P L P -
-----+-----+-----+-----+-----+-----+-----+-----+
16501 GTTGAGGAAAAGGATTCGTTCACTGATTTGTTGGCCACCGTGCAGAACCAAGCCATGTC 16560
-----+-----+-----+-----+-----+-----+-----+-----+
16502 CAACTCCCTTTTCCCTAAGCAAGTCACTAACGACCCCGTGGCACCGTCTTGGTTCGGTACAG

```


Figure 101AM

```

a V E G K G I R S V I A G H R A E P S H V -
    TTTTCCGCCATTGGGATCAGCTAAAAGTTGGAGATGCTCTTTTATTATGATAATGGCCAG + 16620
    -----+-----+-----+-----+-----+-----+-----+
    AAAAAGCCGGTAAACCTAGTCGATTTTCAACCTCTACGAGAAAATAFACTATTACCGGTC
a F F R H L D Q L K V G D A L Y Y D N G Q -
    GAAATTGTAGAATATCAGATGATGGACACAGAGATTTATTTTCCGCGGAATGGGAAAAA + 16680
    -----+-----+-----+-----+-----+-----+-----+
    CTTTAAACATCTTATAGTCTACTACCTGTGTCCTAATRAAAATGGCAGCCCTTACCCTTTT
a E I V E Y Q M M D T E I I L P S E W E K -
    TTAGAAATCGGTTAGCTCTAAAATAATCATGACCTTGTGATAACCTGGATCCGATCCCTACC + 16740
    -----+-----+-----+-----+-----+-----+-----+
    AATCTAGCCAACTCGAGATTTTATAGTACTGGAACHTATTGGACCGTAGGCTAAGGATGG
    L E S V S S K N I M T L I T C D P I P T -
    TTTAATAAACCGCTTATAGTGAATTTTGAACGAGTCGCTGTTTATCAAAAAATCAGATCCA + 16800
    -----+-----+-----+-----+-----+-----+-----+
    AAATTAATTTGGGATTAATCACTTAAAACCTTCTCAGCGACAAAATAGTTTTTAGTCTAGGT
a F N K R L L V N F E R V A V Y Q K S D P -
    CAAACAGCTGCAGTCCGAGGTTGCTTTTACGAAAGAGGACAACTGTATCCCGTGTIT + 16860
    -----+-----+-----+-----+-----+-----+-----+
    GTTTCGACGTC AACGCTCCCAACGAAAATGCTTTCTCTCTGTAGACATAGCGCACAA
a Q T A A V A R V A F T K E G Q S V S R V -
    GCAACTCTCAATGGTTGTACCGTGGCTAGTGGTACTGGCATTTCTGGGAATCCCTGTTT + 16920
    -----+-----+-----+-----+-----+-----+-----+
    CGTTGGAGAGTTACCAACATGGCAGCCGATCACCACCATGACCGTAAAGACCCCTTAGGACAAA
a A T S Q W L Y R G L V L A F L G I L F -
    GTTTTGTGGAAGCTAGCACCTTACTACGAGGAAAATAAAAAGAAATGAAGGAAAAGCTA + 16980
    -----+-----+-----+-----+-----+-----+-----+
  
```

PCT/US05/27239 292/487

Figure 101AN

CAAACACCTTCGATCGTGCAAATGATGCTCCCTTTATTTTTCTTTACTTTCCTTTCGAT

V L W K L A R L L R G K *

AGGCTGTTCCCTTTTCCGGCTCTTTGTCAACTGTAGTGGTGAATAAAGCTAAGCTCG

16981 -----+ 17040

TCCGACAAGGAAAAGGCCGAGAAACAGTTGACATCACCCAACTTTTTTTTCGATTCGAGC

AGAAAGGACAAAATTTGTCCTTTCTTTTTTTTGATATTCAGAGCGATATAAATCCGTTTTTT

17041 -----+ 17100

TCTTTCCCTGTTTAAACAGGAAAGAAAACATAAAGTCTCCCTATTTTTTAGGCATAAAA

GAAGTTTTCAAA

17101 -----+ 17112

CTTCAAAAGTTTT

293/487

M1, strain 2580

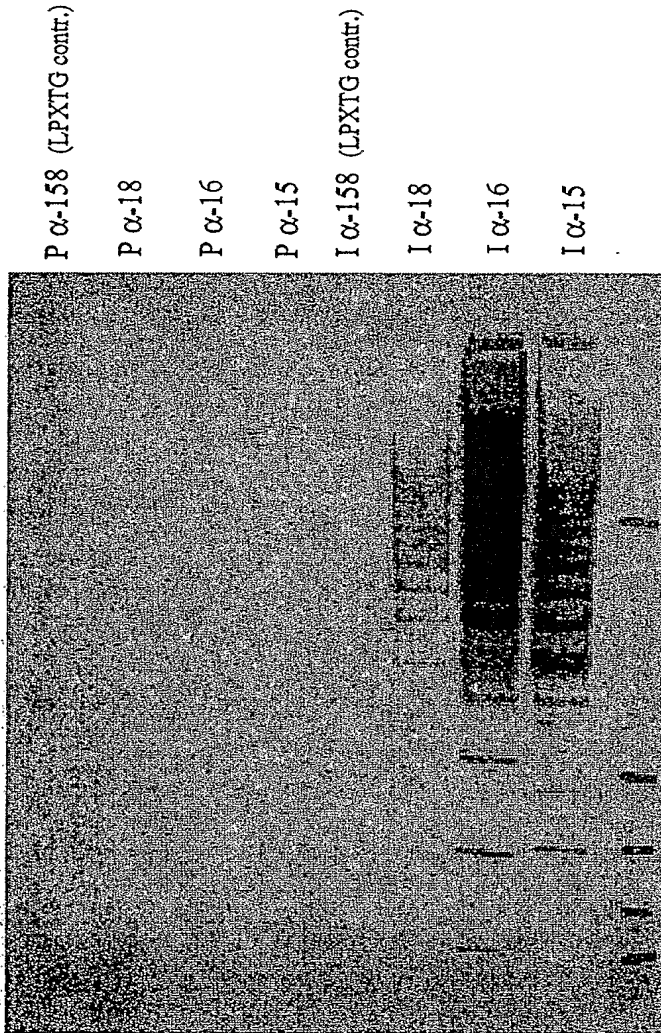


Figure 102

LEGEND:

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

PCT/US05/27239 294/487

M1, strain 2913

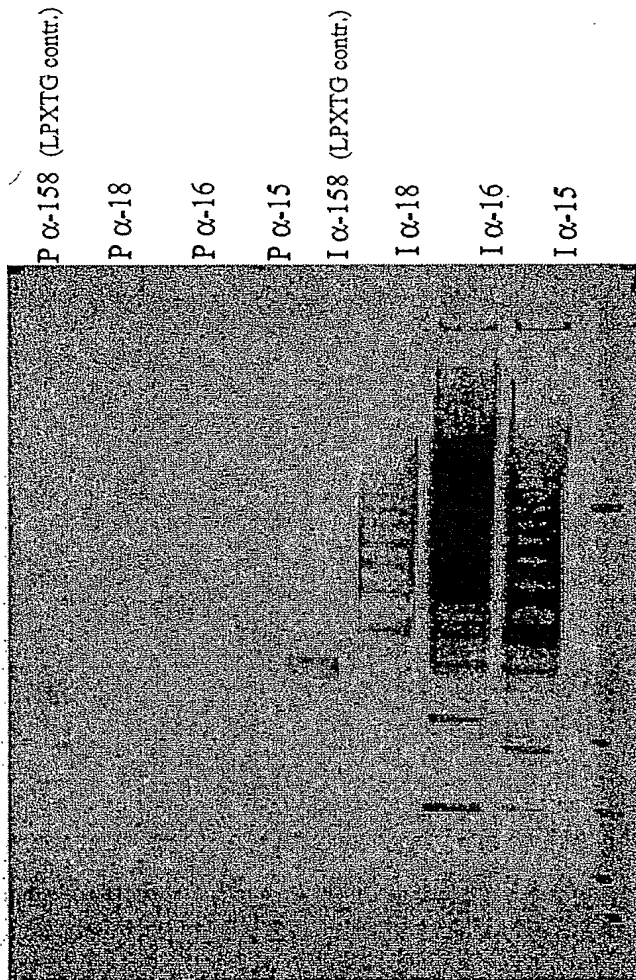


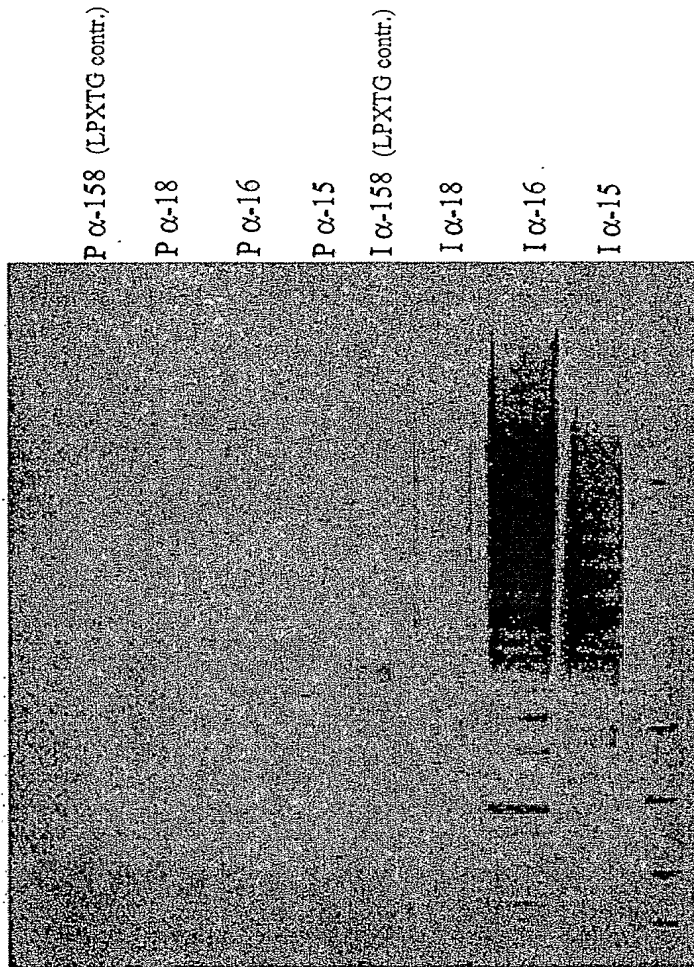
Figure 103

LEGEND:

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

M1, strain 3280



LEGEND:

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

Figure 104

M1 strain 3348

P α -158 (LPXTG contr.)
P α -18
P α -16
P α -15
I α -158 (LPXTG contr.)
I α -18
I α -16
I α -15

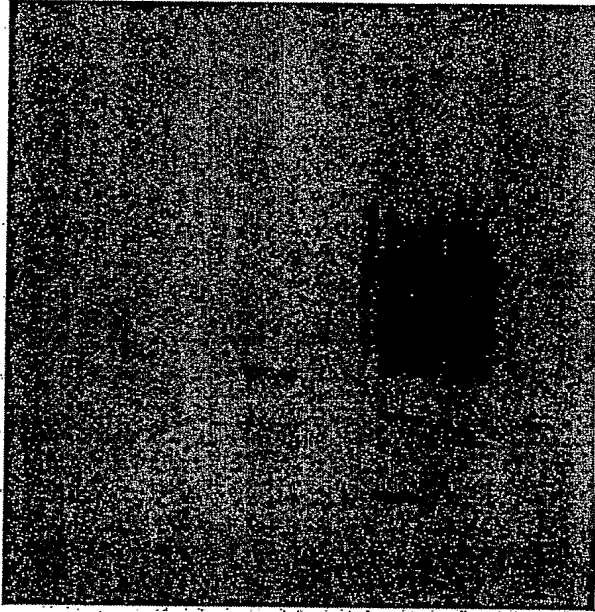


Figure 105

M1 strain 2719

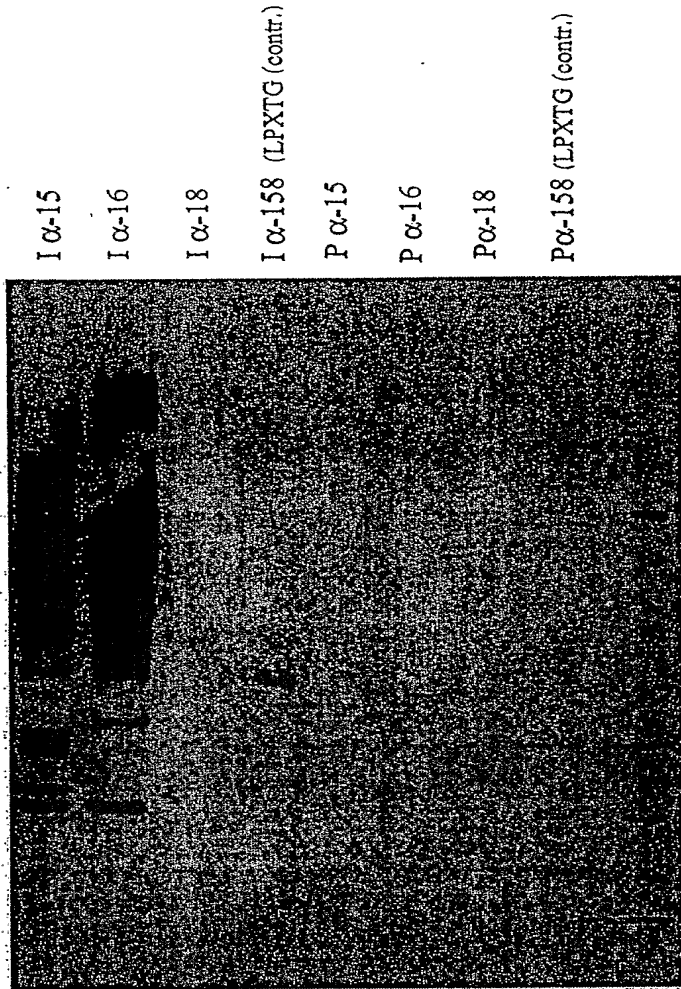


Figure 106

Western blot Western blot on fraction enriched in surface proteins of M1 (SF370)

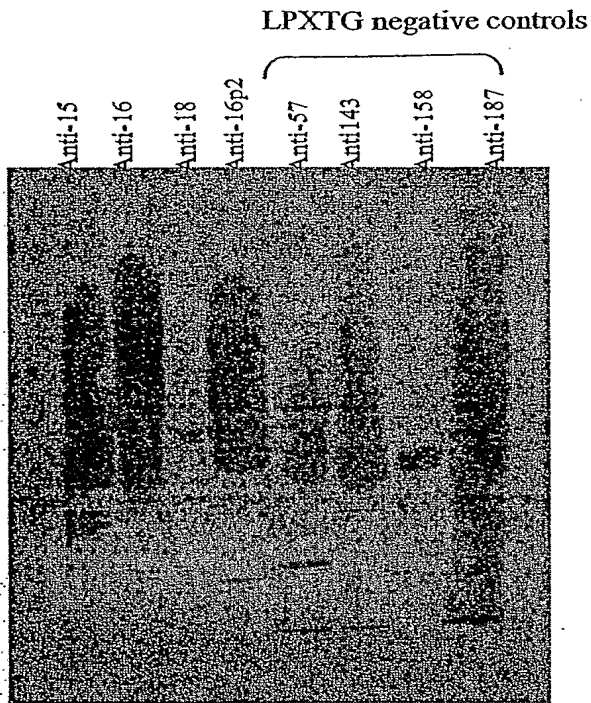
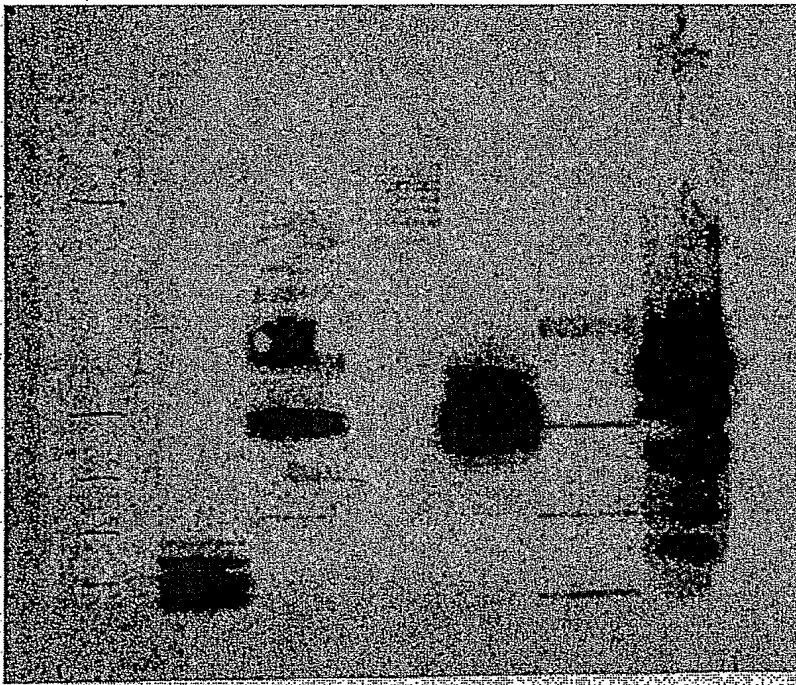


Figure 107

Western blot on fraction enriched in surface proteins of M12 (2728)

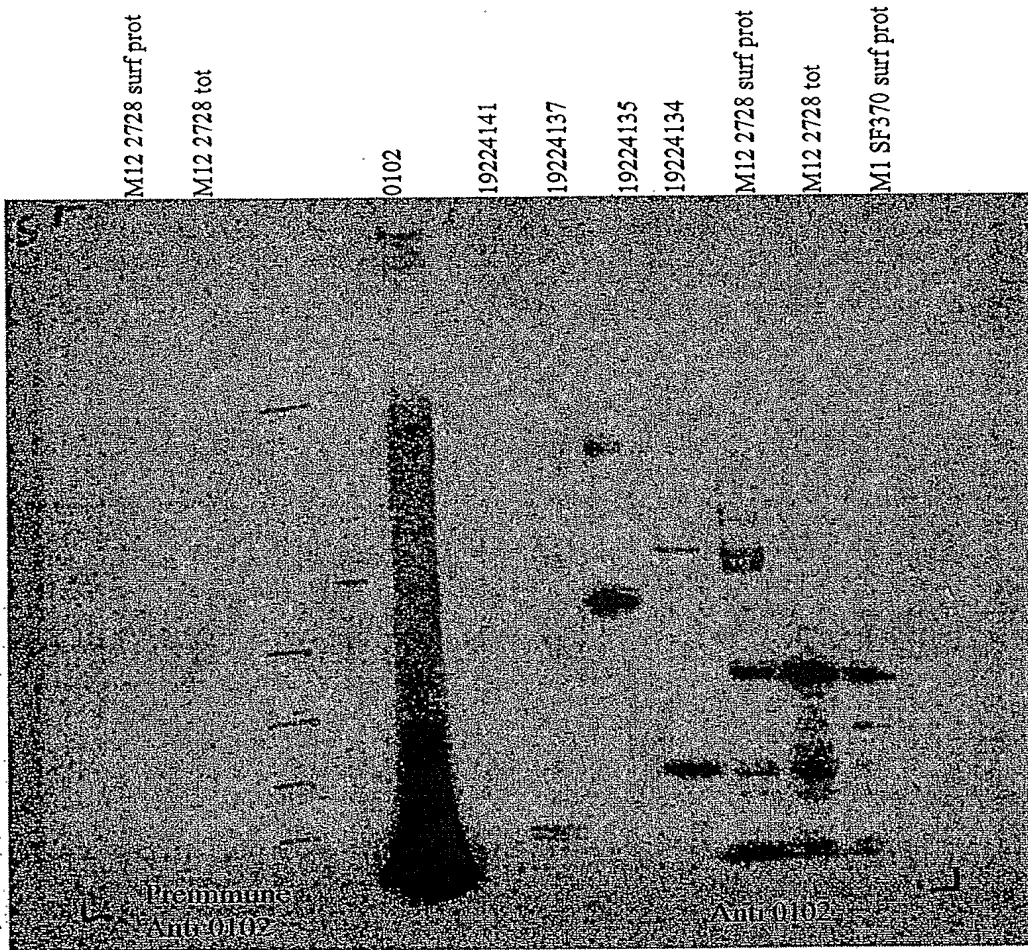
Anti-19224134
Anti-19224135
Anti-19224137
Anti-19224141
Anti-0102
Anti-158 (control)



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2a)", the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in M12 (strain isolate 2728). This means that in M12 three out of the five LPKTG proteins form high-molecular weight structures:

Figure 108

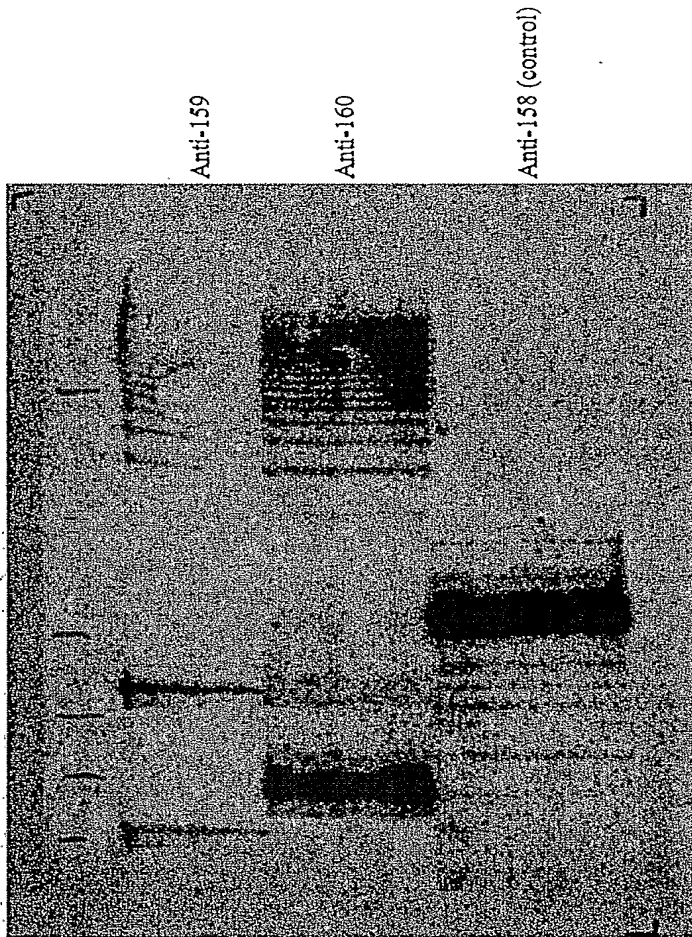
300/487



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2bis)", the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in protein extracts from M12 (strain isolate 2728.)

Figure 109

Western blot on fraction enriched in surface proteins of M6 (2724)

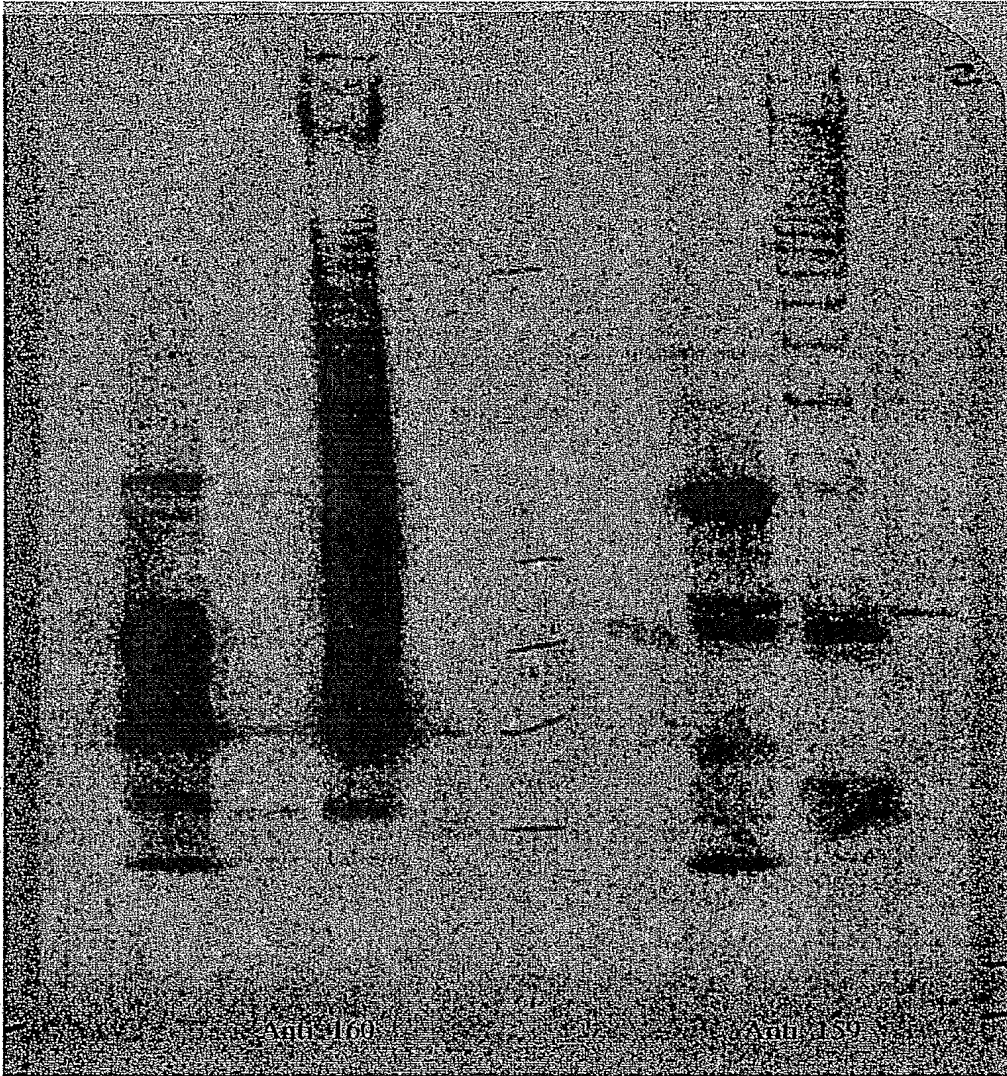


M6 strain isolate 2724

Figure 110

PCT/US05/27239

0160
0159
M6 3650 surf prot
M1 SF370 surf prot
Marker
0160
0159
M6 3650 surf prot
M1 SF370 surf prot



M6 strain isolate 3650

Figure 111

PCT/US05/27239

M6 2724 fraz art

M6 2724 tot

0160

Marker

158

0160

0159

M6 2724 surf prot

M6 2724 tot

M1 SF370 surf prot

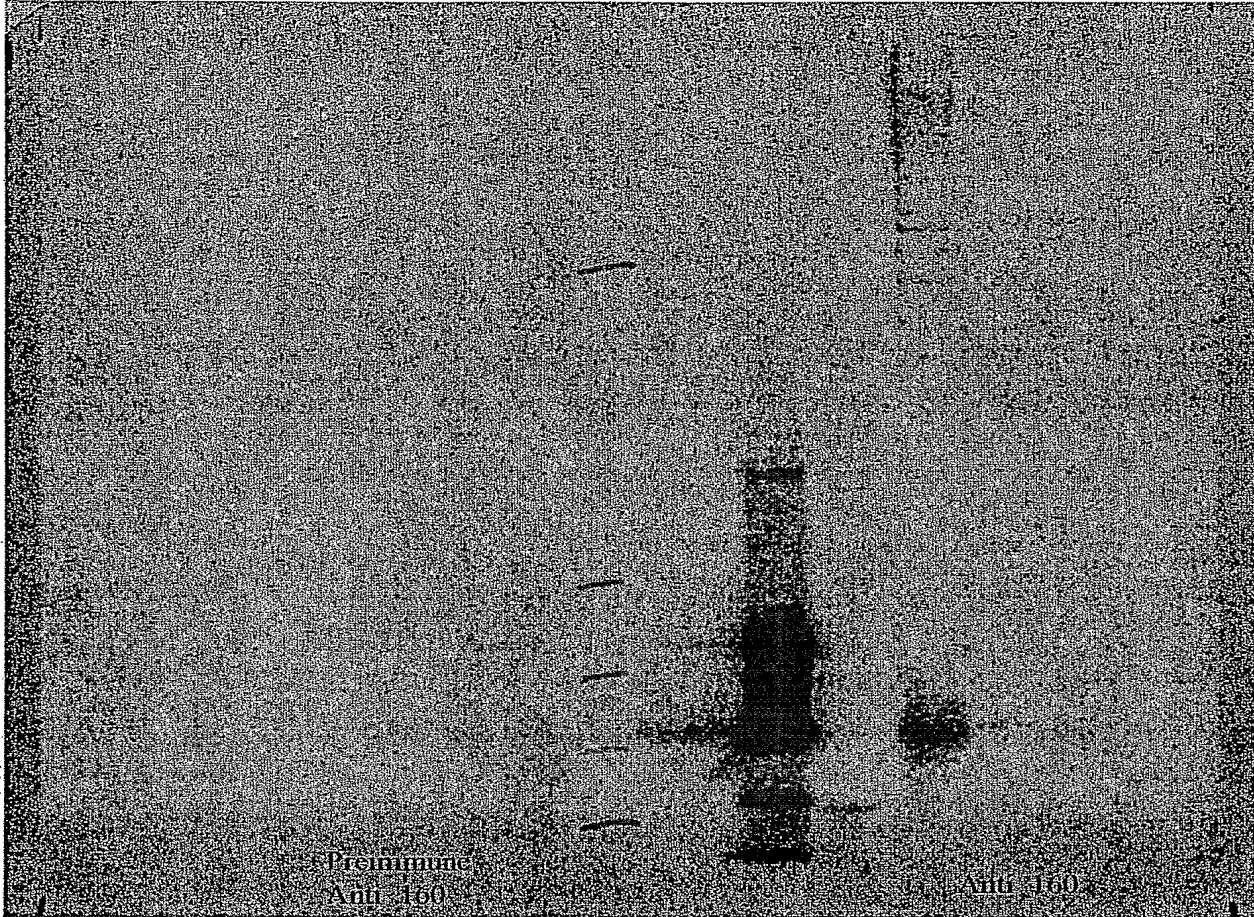


Figure 112

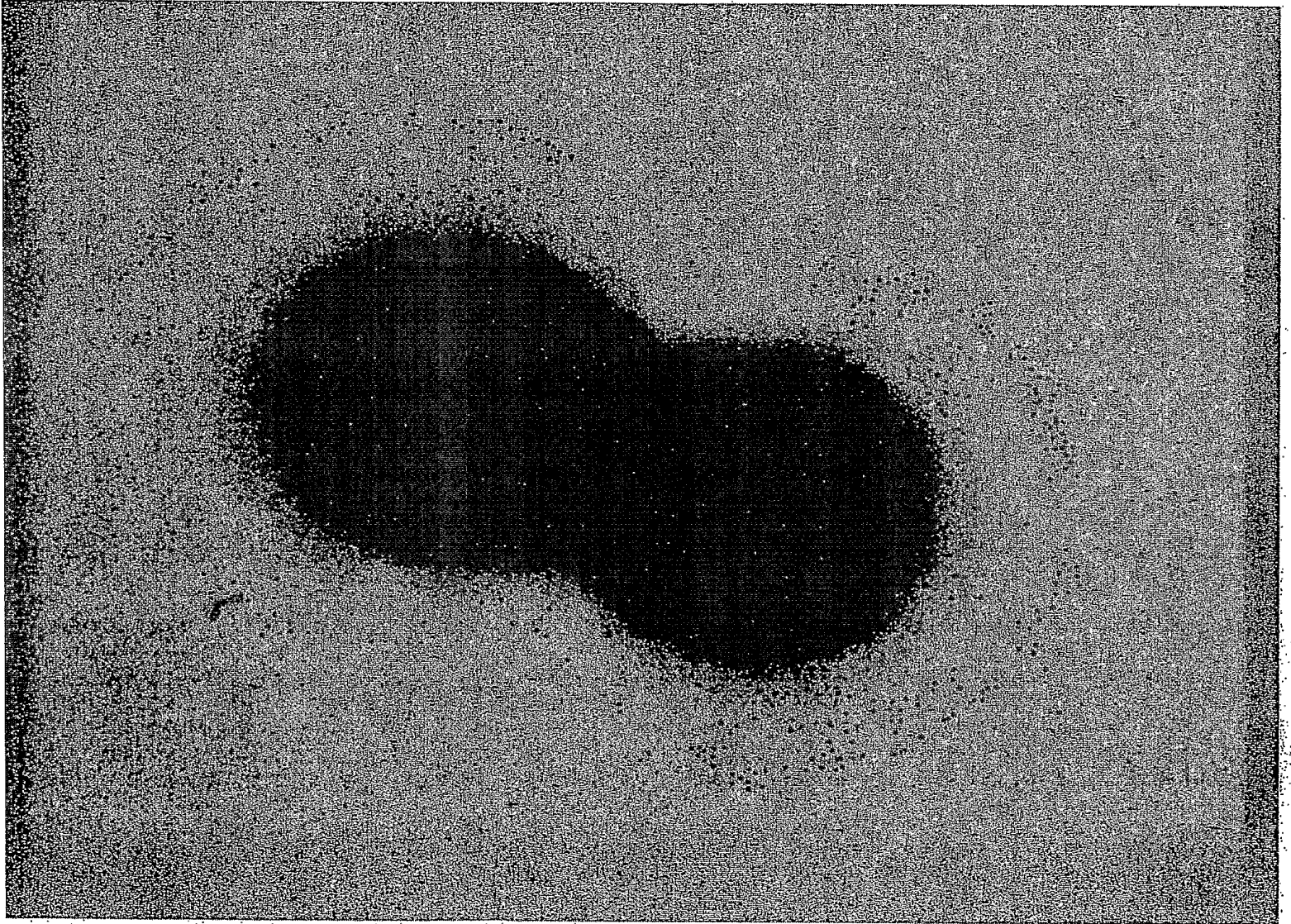


FIGURE 113

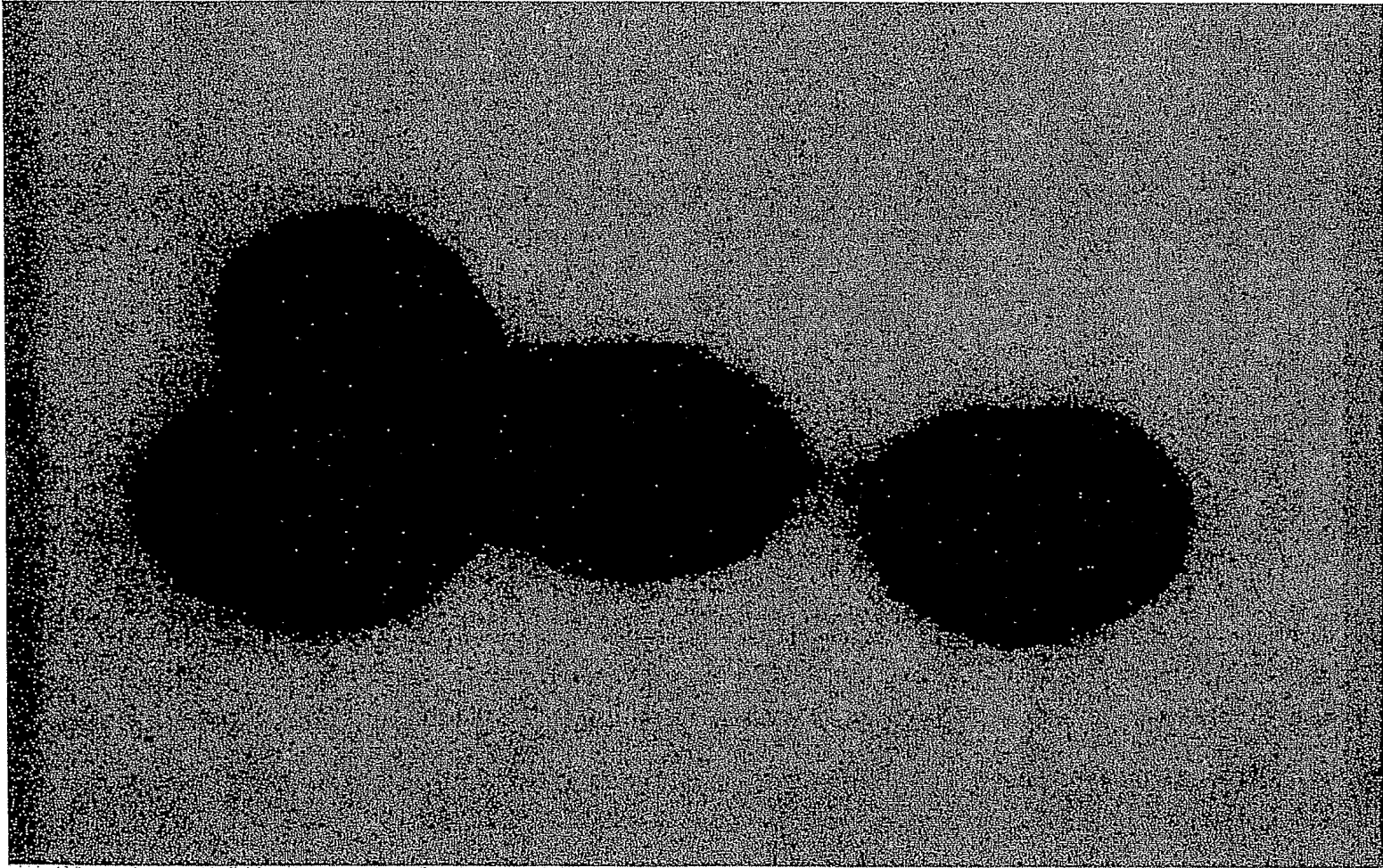


FIGURE 114

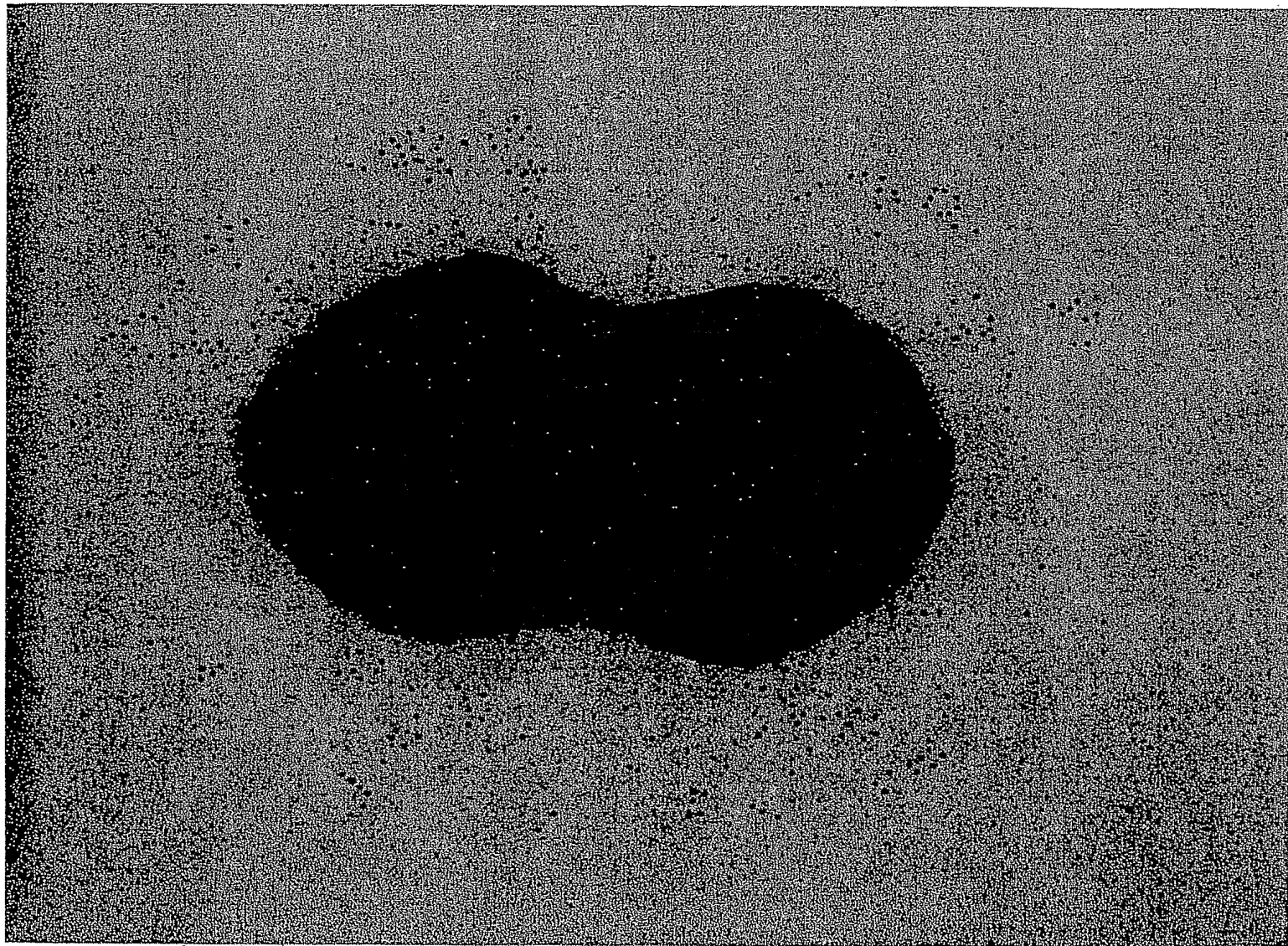


FIGURE 115

PCT/US05/27239 307/487

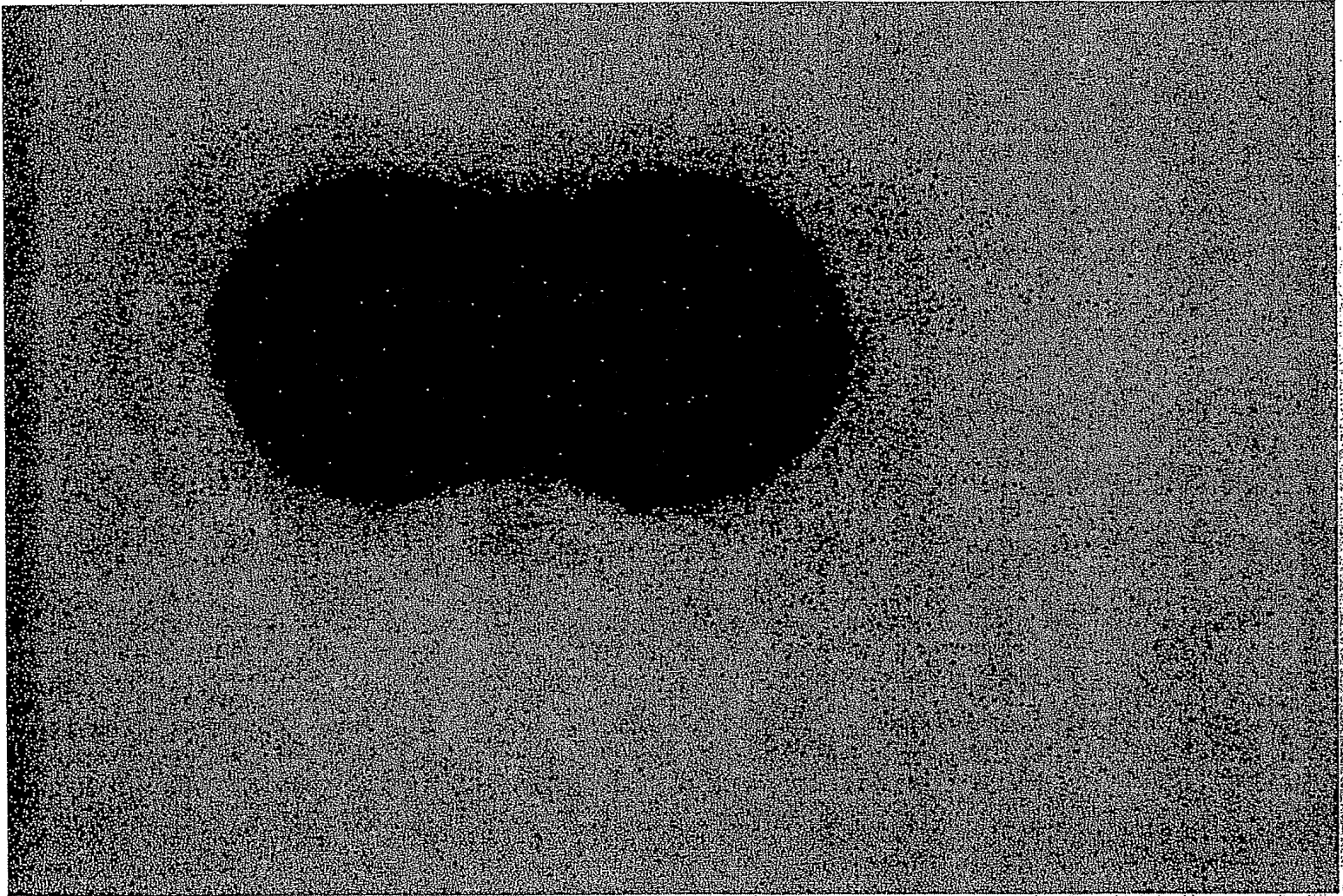


FIGURE 116

PCT/US05/27239 302/487

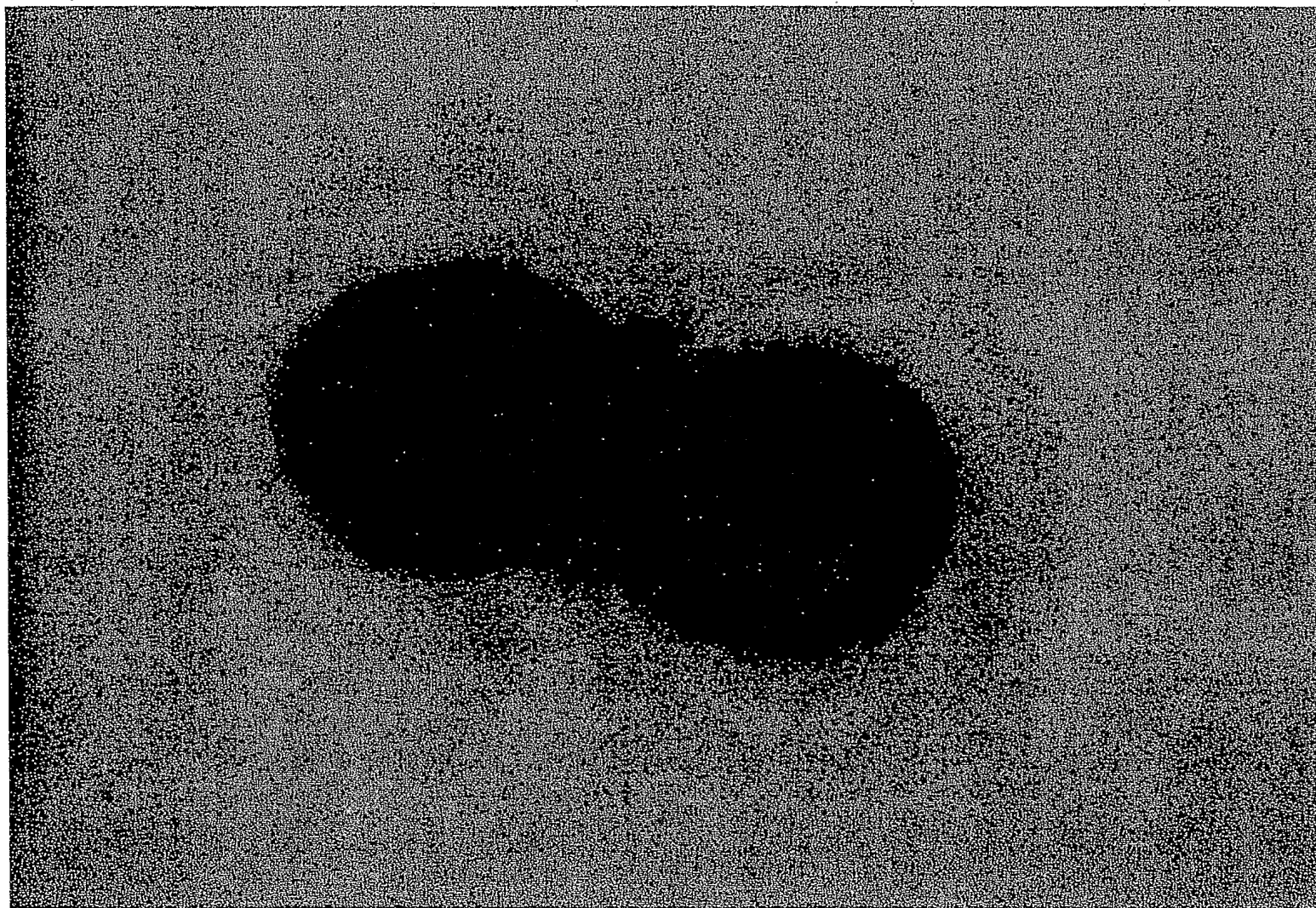


FIGURE 117

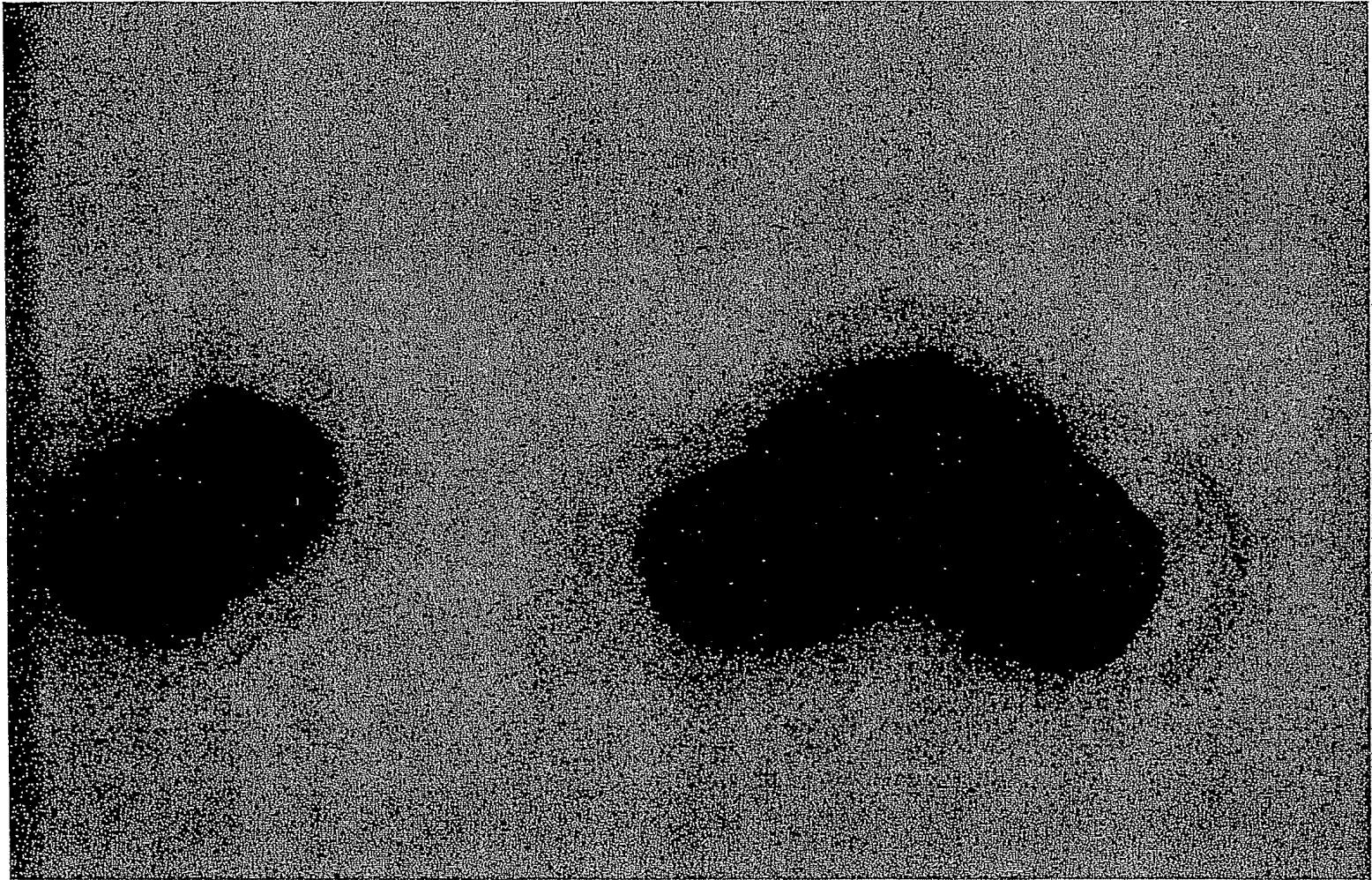


FIGURE 118

PCT/US05/27239 310/487

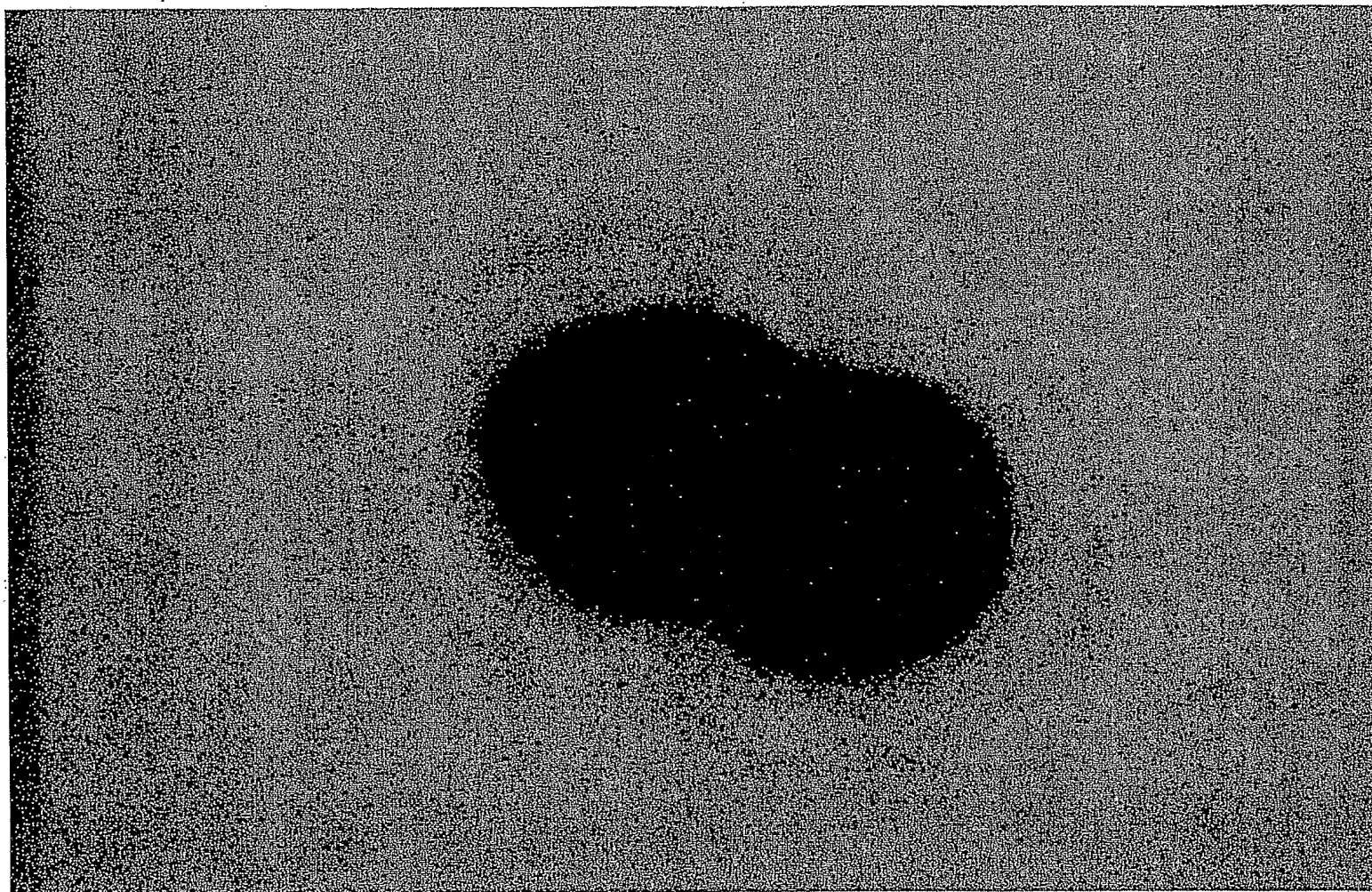


FIGURE 119

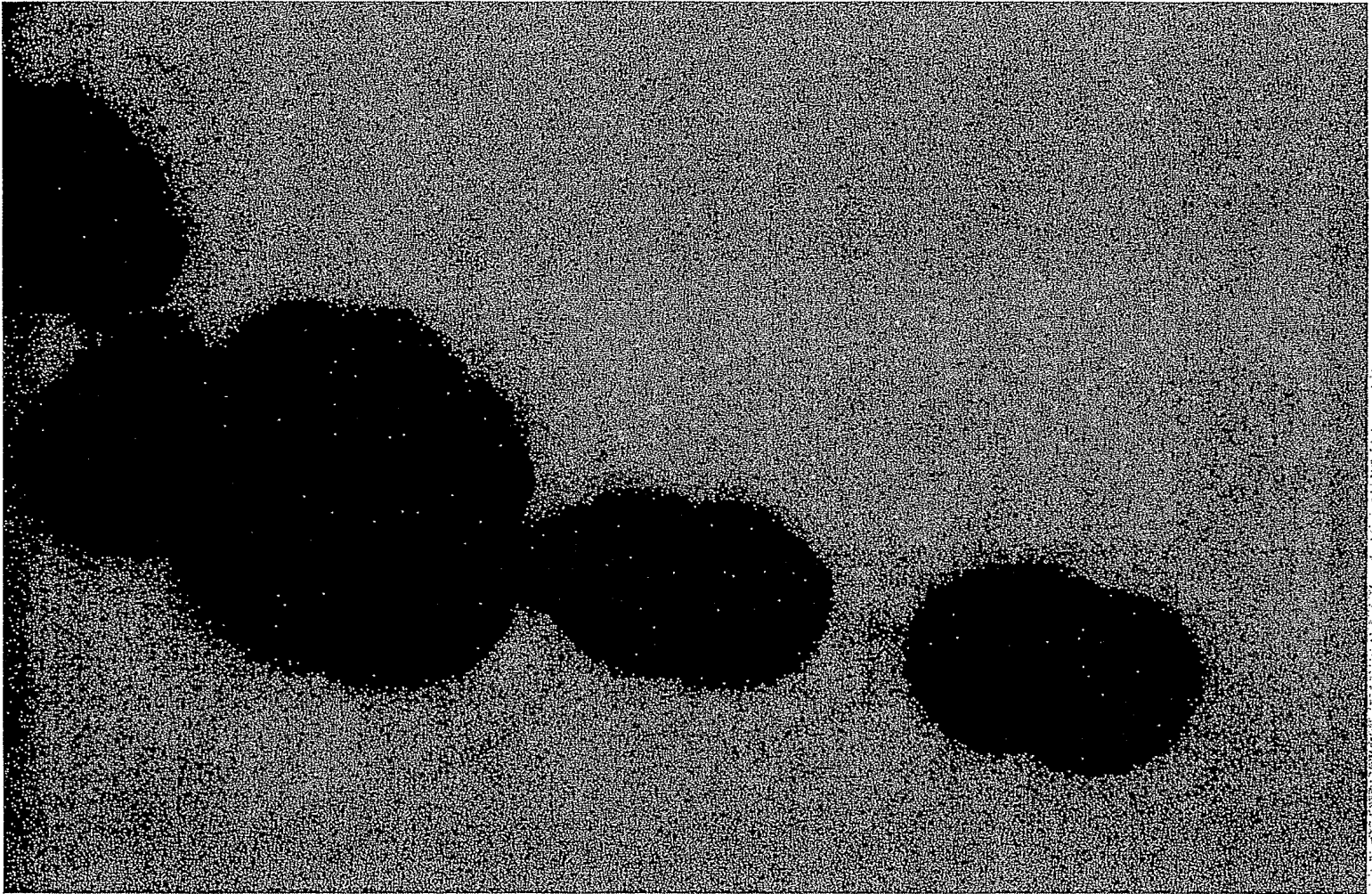


FIGURE 120

312/487

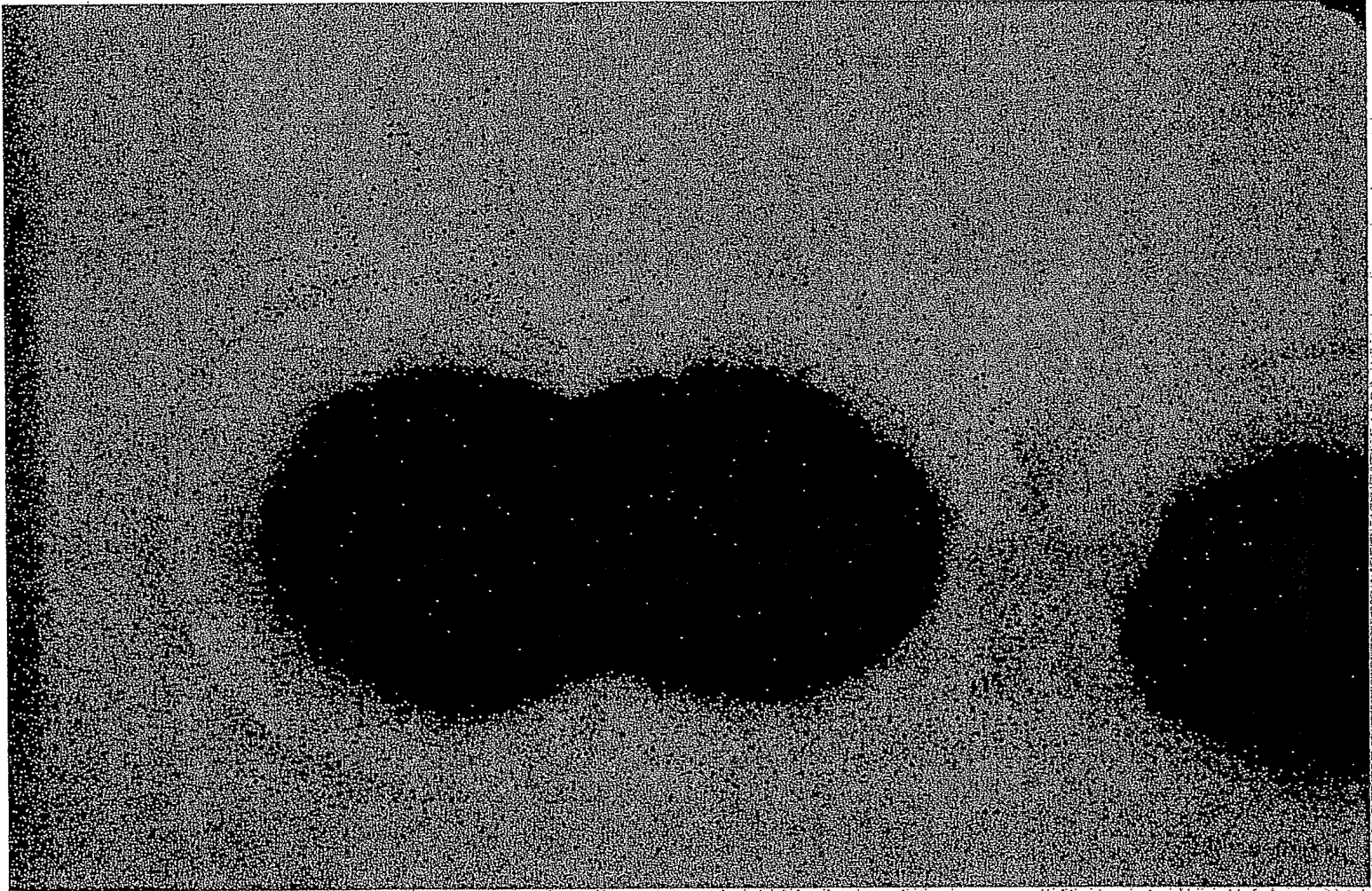


FIGURE 121

313/487

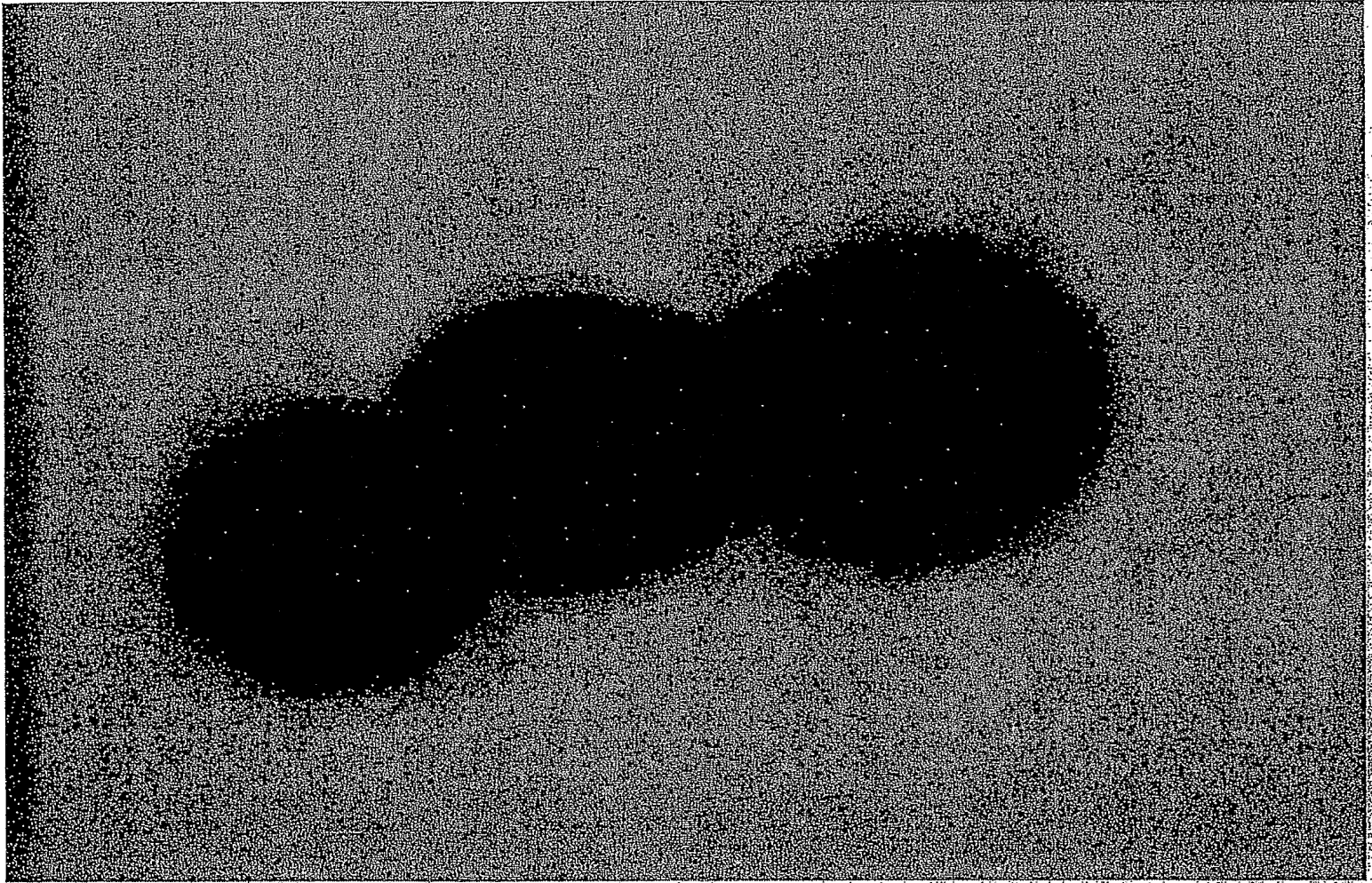


FIGURE 122

314/487

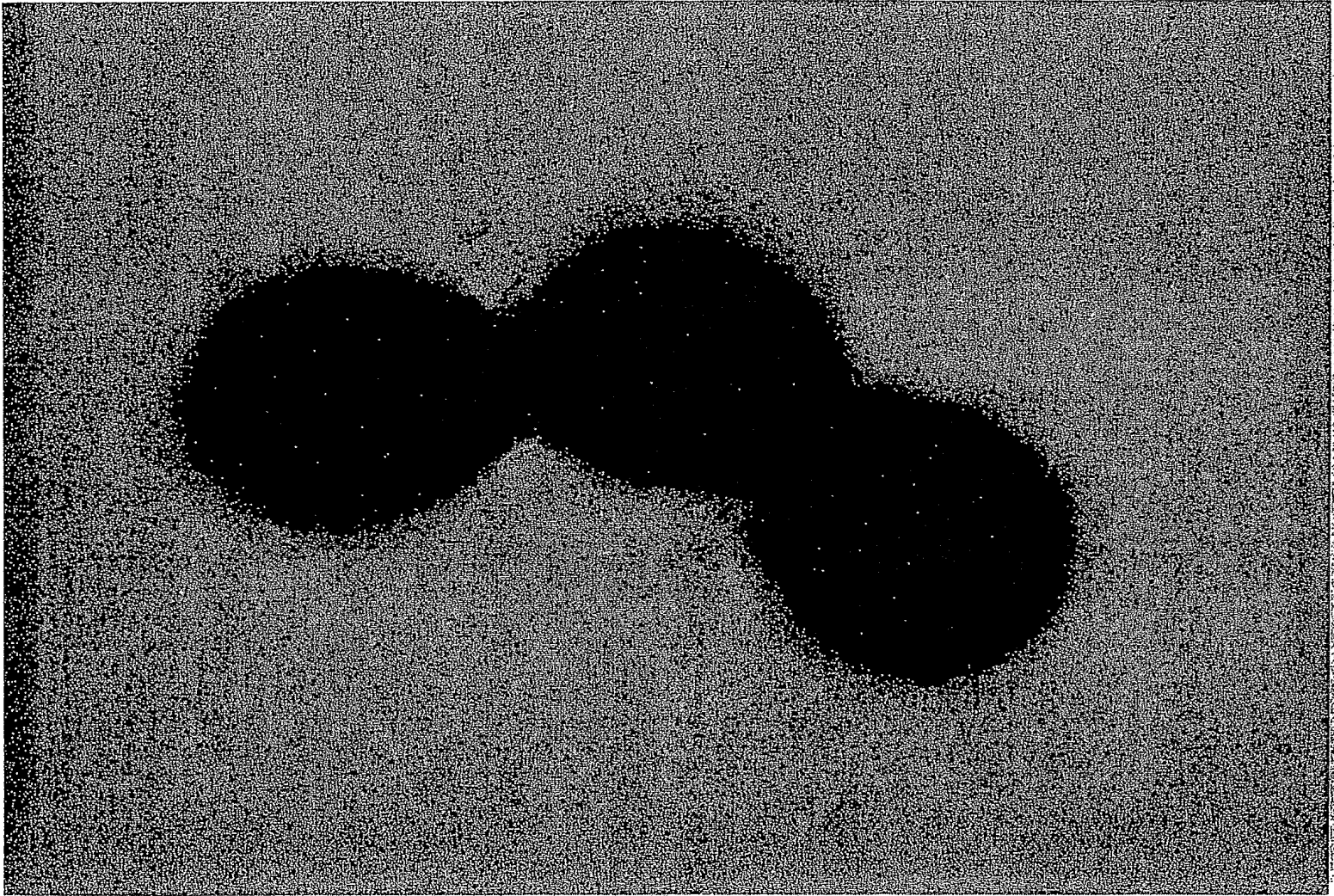


FIGURE 123

315/487

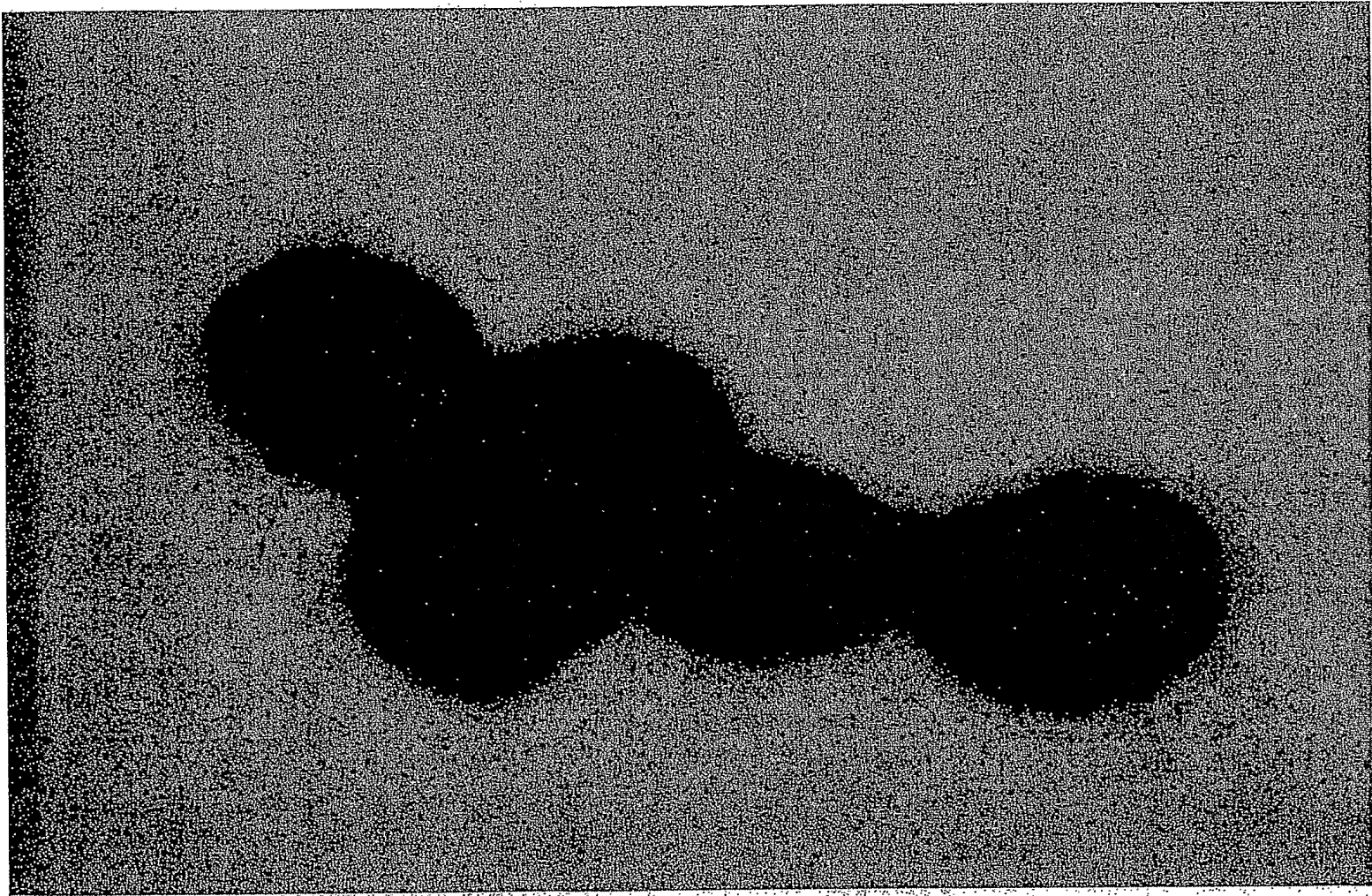


FIGURE 124

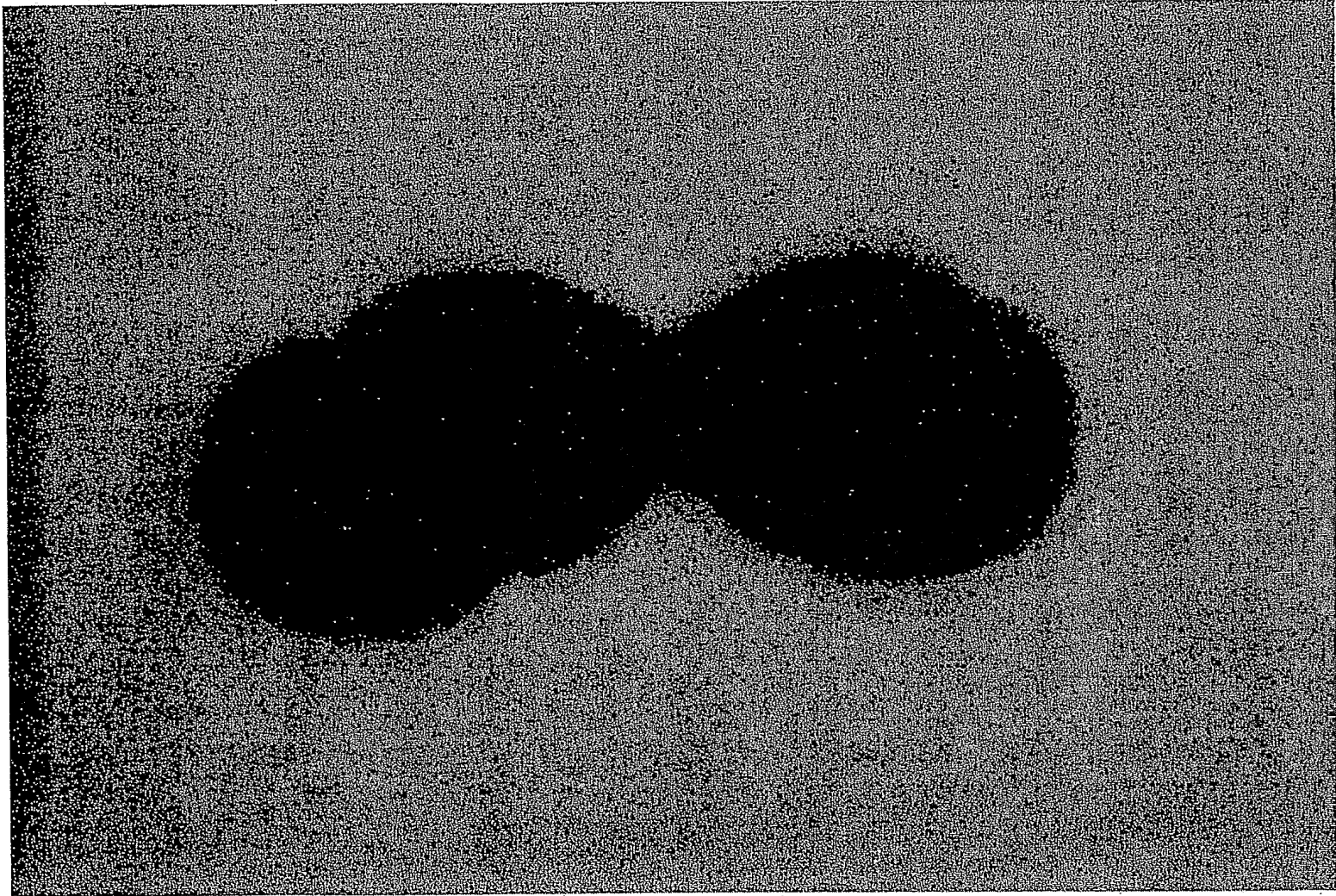


FIGURE 125

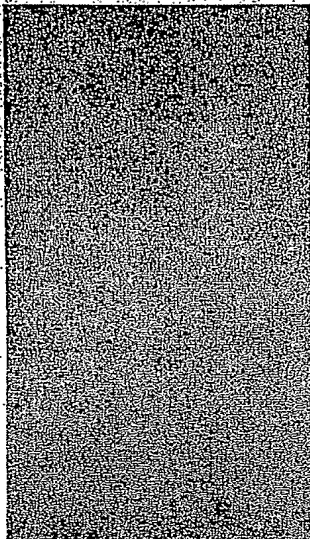
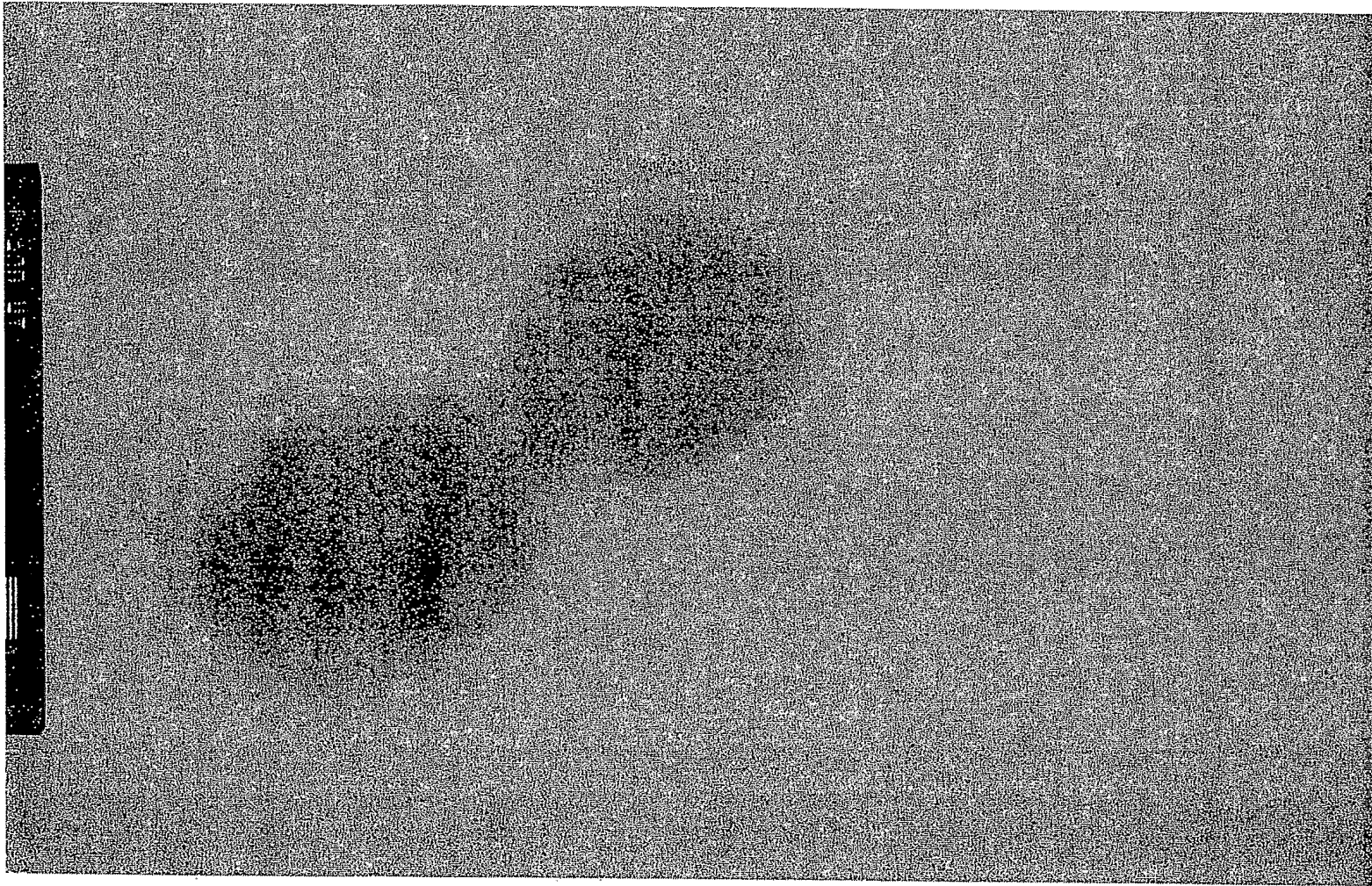


FIGURE 126

Figure 127



PCT/US05/27239 319/487

Figure 128

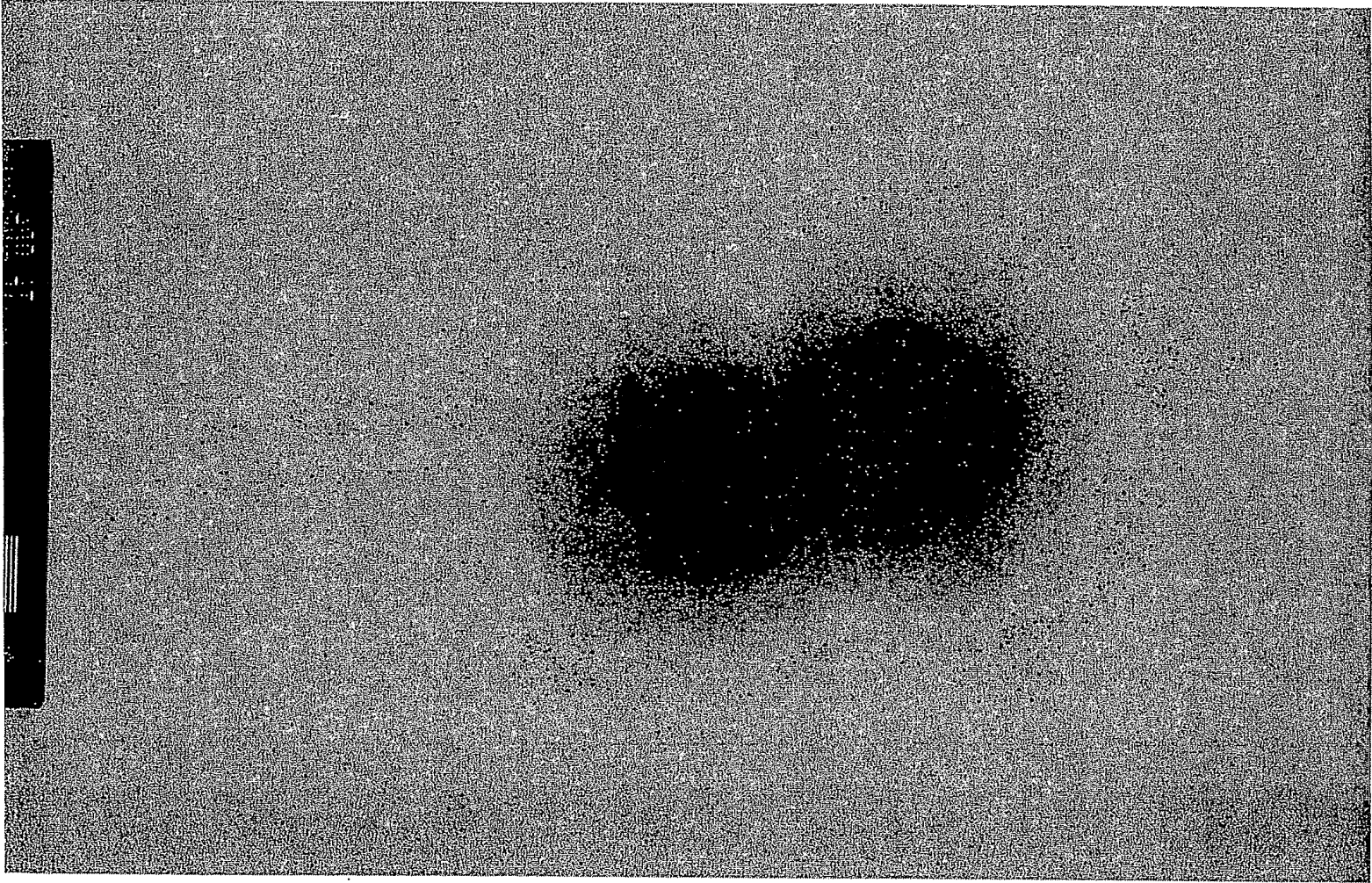


Figure 129

